DOCUMENT RESUME

ED 389 182 FL 023 363

AUTHOR Hopstock, Paul J.; Zehler, Annette M.
TITLE Special Issues Analysis Center (SIAC). Annual Report: Year Three. Volume VI: Task Order D170 Report--Recommendations on Student Outcome Variables for Limited English Proficient (LEP Students). Task Order D190 Report--The Uses of Communications Technology for Language Proficiency Assessment and Academic Assessment.
INSTITUTION Development Associates, Inc., Arlington, Va.
SPONS AGENCY Office of Bilingual Education and Minority Languages Affairs (ED), Washington, DC.
PUB DATE 95
CONTRACT T292001001
NOTE 174p.; The SIAC annual report for year 3 consists of seven volumes, see FL 023 358-364.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; English (Second Language); *Evaluation Criteria; *Information Technology; Language Proficiency; Language Tests; *Limited English Speaking; Minority Groups; Outcomes of Education; *Student Evaluation; *Testing

ABSTRACT

Two reports concerning the evaluation of language minority and limited-English-proficient students are presented. The Task D170 report provides conclusions of a written focus group on the issues involved in defining appropriate outcome variables, outcome variables to be used, and the relative importance of LEP student outcome variables for school accountability and for assessing program effectiveness. Group results are summarized, and the responses of each focus group member on each of five questions are presented. The Task D190 report presents the results of a written focus group on five questions about panelists' experiences with communications technology for this purpose, potential uses, potential for improving the effectiveness and cost effectiveness of assessment, specific technologies holding the most promise, possible difficulties in developing and implementing communications technology for assessment. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FL

ED 389 182

SIAC

Special Issues Analysis Center

Annual Report: Year Three

Volume VI: Task Order 17 Report, Task Order 19 Report (Task Six)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Development Associates, Inc.
Research, Evaluation, and Survey Services Division

FL023343

BEST COPY AVAILABLE

SIAC

Special Issues Analysis Center

Annual Report: Year Three

Volume VI: Task Order 17 Report,
Task Order 19 Report

(Task Six)

1995

Development Associates, Inc.

Research, Evaluation, and Survey Services Division

3

This report was prepared for the U. S. Department of Education, Office of Bilingual Education and Minority Languages Affairs, under Contract No. T292001001, Task No. 6. The opinions, conclusions, and recommendations expressed herein do not necessarily reflect the position or policy of the Department of Education and no official endorsement by the Department of Education should be inferred.

SIAC

SPECIAL ISSUES ANALYSIS CENTER

DEVELOPMENT ASSOCIATES, INC.
1730 NORTH LYNN STREET, ARLINGTON, VIRGINIA 22209-2023
TEL: (703) 276-0677 FAX: (703) 276-0432

YEAR THREE ANNUAL REPORT

The Special Issues Analysis Center (SIAC), as a technical support center, provides assistance to the Office of Bilingual Education and Minority Languages Affairs (OBEMLA), U.S. Department of Education (ED). The purpose of the SIAC is to support OBEMLA in carrying out its mission to serve the needs of limited English proficient students. In this role, the SIAC carries out data analysis, research, and other assistance to inform OBEMLA decision-making. These activities are authorized under the Bilingual Education Act of 1988, Public Law 100-297.

The responsibilities of the SIAC are comprised of a variety of tasks. These tasks include data entry and database development, data analysis and reporting, database management design, design of project accountability systems, and policy-related research and special issues papers. This report describes activities carried out by the SIAC in Year Three. A full list of SIAC products for all three years of operation is presented in the Appendix.

This Annual Report consists of seven volumes, which include the overview report on the SIAC activities in Year Three plus six additional volumes. These volumes present copies of selected reports submitted to OBEMLA by the SIAC in the past year, including copies of all task order reports submitted. The contents of each volume are outlined below:

Volume I: Overview of SIAC activities in Year Three;

Volume II: Copies of Short Turnaround Reports (STRs) based on analyses of Title VII application data and other data related to LEP students;

Volume III: The SEA Report/Task Seven;

Volume IV: Task Order 12 and Task Order 13 Reports;

Volume V: Task Order 10 and Task Order 16 Reports;

Volume VI: Task Order 17 and Task Order 19 Reports; and,

Volume VII: Task Order 16 and Task Order 21 Reports.

SIAC *Special Issues Analysis Center*

Recommendations on Student Outcome Variables for Limited English Proficient (LEP) Students

Task Order D170
Written Focus Group Report

May 5, 1995

Development Associates, Inc.
Research, Evaluation, and Survey Services Division

This report was prepared for the U. S. Department of Education, Office of Bilingual Education and Minority Languages Affairs, under Contract No. T292001001, Task Order No. D170. The opinions, conclusions, and recommendations expressed herein do not necessarily reflect the position or policy of the Department of Education and no official endorsement by the Department of Education should be inferred.

Prepared by:

Paul J. Hopstock

Development Associates, Inc.

Prepared for:

**Office of Bilingual Education and
Minority Languages Affairs
U. S. Department of Education**

TABLE OF CONTENTS

I. INTRODUCTION 1

II. ABSTRACT 2

III. FINDINGS 4

A. Issues in Defining Appropriate Outcome Variables 4

1. To what extent should common versus site-specific outcome variables and measures be used in schools of different types and with different objectives? 4

2. Should analysis of outcome measures for LEP students take into account the opportunity to learn (e.g. based on the courses which were provided and the nature of the material in those courses)? 5

B. Outcome Variables To Be Used With LEP Students 5

1. Academic Achievement 6

2. Language Proficiency 10

3. Behavioral Outcomes Related to Student Effort 13

4. Psychological Outcomes 15

5. Work Readiness Outcomes 17

C. The Relative Importance of LEP Student Outcomes for School Accountability 18

1. Elementary Schools 18

2. High Schools 19

IV. CONCLUSIONS AND RECOMMENDATIONS 20

A. Issues in Selecting Outcome Measures 20

1. What are the purposes of assessment? 20

2. Should native language development and the inclusion of examples using LEP student cultural backgrounds be goals in the instruction of LEP students? 21

3. What language(s) should be used for assessment? 21

4. Should measures of growth be used for LEP students instead of norms or criteria developed for mainstream students? 22

B. Recommendations Concerning LEP Student Outcome Measures 23

1. Recommendations to Researchers and Evaluators 23

2. Recommendations to OBEMLA 24

APPENDICES

Appendix A: List of Focus Group Participants

Appendix B: Focus Group Questions

Appendix C: Written Recommendations from the Participants

I. INTRODUCTION

A written focus group on outcome variables for limited English Proficient (LEP) students was coordinated by the Special Issues Analysis Center of Development Associates, Inc. of Arlington, Virginia in February and March of 1995. The purpose of the written focus group was to identify the most pertinent LEP student outcomes for schools serving LEP students and which are undergoing school reform, and to provide detail concerning the measurement, administration, and analysis of those outcomes. The information was intended to assist the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) of the U.S. Department of Education in fulfilling its mission to provide national leadership in promoting equal access to high quality education for language minority populations. OBEMLA was particularly interested in outcome variables which would generate findings with evaluative and policy implications, and which could be used either in national studies including LEP students or in systematic accountability assessments.

Four researchers who have been actively involved in issues relating to LEP student assessment participated in the written focus group. Three of the participants were university researchers with active research interests in student assessment and the education of LEP students. The fourth participant was an educational evaluator with extensive experience studying programs for LEP students at the local school system level. These participants were sent a list of five questions which they were asked to address, and were given approximately four weeks to provide written responses.

This report presents the results of the written focus group. The comments of the panelists are summarized in the Findings chapter of this report. The Findings chapter includes three major sections:

A. Issues in Defining Appropriate Outcome Variables
B. Outcome Variables To Be Used With LEP Students
C. The Relative Importance of LEP Student Outcome Variables for School Accountability

The report also includes the recommendations of Development Associates, Inc. concerning LEP student outcomes based on the panelists' comments. There are three appendices: Appendix A provides a list of the panelists and their affiliations; Appendix B presents the questions as they were provided to the panelists; and Appendix C provides the panelists' written answers to the questions organized in the same way as the Findings chapter (see Sections A-C above).

II. ABSTRACT

The written focus group was organized aro· nd five questions which were sent to panelists. Shortened versions of the questions and summary answers to them are presented below.

In schools serving LEP students and which are undergoing school reform, what are the most pertinent LEP student outcomes that should be examined when considering the impact of such school reforms?

Panelists were given five categories of LEP student outcome variables, and were asked to list variables within those categories: (a) academic achievement in core subject areas; (b) language acquisition; (c) behavioral variables indicating student effort or motivation; (d) psychological variables; and (e) readiness for the world of work. Panelists provided a wide range of variables within each of these categories. Panelists made three types of recommendations about the selection of variables to assess school reform: (1) the variables and measures which are used should relate to the specific objectives of the school; (2) a range of outcome variables should be used to gain a comprehensive picture of school effects; and (3) LEP students' opportunity to learn (exposure to challenging content, etc.) should be examined along with LEP student outcomes.

Please select three or four specific LEP student outcomes, and describe how you would operationalize and measure the outcomes. Describe the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures.

Panelists focused most of their attention on variables relating to achievement in core academic areas and language proficiency. To measure achievement in core academic areas, they pointed to some existing assessments as models (NAEP, the New Standards Project, the California Learning Assessment System, the New York Program Evaluation Test in Science), but they generally believed that new assessments should be developed to measure a broader range of concepts at more grade levels. Panelists strongly supported performance assessments and portfolio assessments as measures of academic achievement, and suggested a number of approaches which have been used for such assessments. They cautioned, however, that care should be taken in designing and validating such measures for use with LEP students. In the area of language proficiency, panelists mentioned both the LAS and LAB as existing measures which could be used for assessments. However, they pointed to the limitations of these measures for assessing particular skills and particular levels of language development, and also suggested that performance assessments and portfolio assessments be used for these purposes.

For the same LEP student outcomes, please describe the appropriate assessment procedures and schedules for assessment. This would include who should be assessed, when and how often they should be assessed, and what special persons, resources, and/or staff training are required for the assessment.

Panelists indicated that appropriate assessment procedures depend upon the purposes of assessment. In general, if the assessment is for broad-scale accountability purposes, panelists were more likely to suggest sampling of students within particular grade levels. If the purpose is for program evaluation or if one purpose is student placement, then panelists were less likely to suggest sampling. Some of the academic achievement measures which were suggested would involve assessments of students at only selected grade levels. Assessments in most of the other areas involve testing or data collection at all grade levels. Panelists indicated that significant training and support would need to be provided for teachers if performance assessments and portfolio assessments are used.

For the same LEP student outcomes, please indicate how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms. What comparisons should be made, and what standards should be used for assessing effectiveness?

Panelists made three key points on the issue of standards and comparisons: (1) the objectives for LEP students should be to meet the same challenging academic standards as for all other students; (2) the evaluation of outcomes for LEP students should take into account their previous educational backgrounds, educational experiences (i.e., opportunity to learn), and language proficiency levels in the languages of instruction and assessment; and (3) for LEP students with limited educational backgrounds or very limited language proficiency in the languages of instruction and/or assessment, standards relating to change or growth should be used rather than standards or criterion scores developed for mainstream students. Panelists suggested a number of approaches for implementing these suggestions.

If you were to hold a school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? Please justify your choices and weighting.

Some of the panelists resisted answering this question because they believed that schools should define their own accountability systems based on their unique objectives. Among the panelists who responded, there was a strong emphasis on growth in academic achievement in core subject areas and increasing mastery of both English and the native language. At the secondary level, there was also a strong emphasis on readiness for post-secondary instruction and for the world of work.

III. FINDINGS

The results of the written focus group are presented in three major sections. In the first section, we discuss two generic issues identified by panelists as being related to the selection and use of outcome measures for LEP students. In the second section, we summarize the comments of panelists concerning the selection, measurement, and use of specific outcome measures. In the third section, we describe the panelists' comments concerning the relative importance of the various outcome measures in elementary and high school settings.

A. Issues in Defining Appropriate Outcome Variables

In defining appropriate outcome variables for LEP students, panelists identified two major issues which affect choices (1) To what extent should common versus site-specific outcome variables and measures be used in schools of different types and with different objectives?; and (2) Should analysis of outcome measures for LEP students take into account the opportunity to learn (e.g., based on the courses which were provided and the nature of the material in those courses)? In this section, we briefly describe those issues. These issues also serve as themes which recur throughout our descriptions of the panelists' comments.

1. To what extent should common versus site-specific outcome variables and measures be used in schools of different types and with different objectives?

Panelists suggested that before it is possible to define appropriate outcome measures for LEP students, it is necessary to describe the goals of instruction for those students. Some of the panelists expressed reservations about applying specific student outcomes as national standards. One of the reasons for those reservations was that panelists did not believe that schools should be held accountable for outcomes which were not relevant to their goals.

One panelist proposed four major factors which would affect the relevance of particular academic achievement outcomes for a school: (1) the grade levels served by the school (e.g., high schools have more diverse objectives for students than do elementary schools); (2) the mission of the school (e.g., schools employing cross-disciplinary curricula have distinct objectives); (3) whether or not schools are organized departmentally (e.g., departmentalized schools have more of a content focus, while non-departmentalized schools have more of a student focus); and (4) the particular reform focus of a school (e.g., some schools are focusing on specific student outcomes as their reform focus). In his words: "...schools and school personnel will interpret the goals for LEP students in terms of their dominant world views. These world views will vary as a function of, at least, the above four characteristics of the schools. And in some cases, quite frankly, these recommendations [the ones made in his paper] will simply be wrong from the perspective of the school in question."

2. **Should analysis of outcome measures for LEP students take into account the opportunity to learn (e.g., based on the courses which were provided and the nature of the material in those courses)?**

A related issue concerns whether the content which LEP students are offered in their classes should be taken into account in developing standards for their academic achievement outcomes. LEP students are often offered classes with less challenging academic content or with less effective modes of academic instruction. One of our panelists thus suggested that all outcome measures of academic achievement be accompanied by measures of "opportunity-to-learn."

Two of the panelists listed opportunity-to-learn variables as important mediators of other LEP student outcomes. One panelist, for example, defined opportunity to learn in core subject areas as follows: "LEP students will have access to and participate in the full mathematics, social studies/geography, and science curricula. In other words, they will have access to challenging content -- the same content that should be made available to all students." She also listed four opportunity-to-learn variables (objectives) related to achievement in core subject areas:

(1) In schools with departmentalized instruction, LEP students will participate in classes with demanding academic content in proportions at least similar to if not higher than non-LEP students.

(2) The language(s) and materials used for instruction in these subjects will be linguistically appropriate to the needs of LEP students.

(3) The students' native language and cultures will be positively reflected in classroom activities and the school climate.

(4) The needs of LEP students are consistently considered in school-wide academic planning and decision making.

The same panelist defined opportunity to learn within the area of language acquisition as follows: "Students must be provided with ESL instruction sufficiently differentiated to meet a full range of student needs. This instruction should support the student through the development of high-order oral communication and literacy skills."

B. Outcome Variables To Be Used With LEP Students

Panelists were asked to discuss LEP student outcome variables within five categories: (1) academic achievement outcomes; (2) language proficiency outcomes; (3) behavioral outcomes indicating student effort or motivation; (4) psychological outcomes; and (5) work readiness outcomes. This section is organized around those categories. Within each of those categories, we present the panelists' comments concerning: (a) what outcome variables

should be studied; (b) how those outcome variables should be measured; (3) how the data on those outcome measures should be collected; and (4) how the data from outcome measures should be analyzed.

1. Academic Achievement

For the purpose of this task order, we have defined academic achievement outcomes as those relating to core subject areas excluding language proficiency. Generally this was meant to focus on mathematics, science, and social studies, though panelists did mention other subject areas such as fine arts, health and physical education, and vocational/technical education. In the sections which follow, we summarize the panelists' recommendations concerning which outcome variables should be used for LEP students and their specifications about how those outcome variables should be measured.

a. Key Outcome Variables To Be Studied

Panelists generally agreed that it is important to examine LEP students' mastery of challenging content in the areas of mathematics, science, and social studies. They used somewhat different ways of describing what they meant by mastery of challenging content, however. One panelist, for example, wrote that measures should "assess student's conceptual understanding, relevant prior knowledge, and ability to apply methods (and discourse) of the discipline." A second panelist wrote of linking assessments to performance standards "that include attention to lower level basic skills and conceptual understanding, problem solving, and conceptual application and communication in a subject matter area. A third panelist defined mastery as "literacy" in the content area. He defined this term as follows:

> "By literacy I mean that an individual has some familiarity with a particular domain so that when he or she encounters a significant and realistic problem requiring knowledge in that domain, that person can make some sense of the problem, can use her or his knowledge of that domain to (a) generate new knowledge, or (b) figure out a way of solving the problem, (c) find someone else who can solve the problem, and (d) understand how the solution fits the problem at hand. In my conception of literacy, detailed technical knowledge of a domain is not required; but understanding of some of the central ideas and how they are inter-related among one another and to specific situations is required."

This panelist went on to write that "...it would seem desirable to strike a balance between broad knowledge on the one hand and in-depth knowledge of 1 or 2 areas... This, of course is a variation of the argument for coverage of a core curriculum plus student choice to focus on those areas that are of interest."

Although academic achievement outcomes were the main topics discussed by panelists, one panelist added one additional variable related to core academic areas: the ability to self-assess academic outcomes. The panelist described this in terms of the "acquisition of metacognitive abilities." Under this formulation, students would be able to take a broader

view of the learning process and would be better able to monitor and adjust their own learning.

b. Measurement of Outcome Variables

The four panelists all provided perspectives on the measurement of academic achievement in core content areas. One panelist discussed achievement in mathematics, one discussed achievement in science, and the other two described more generic approaches to measuring achievement in core content areas.

The panelist who specifically discussed mathematics achievement proposed that outcome measures include both complex performance tasks and portfolio assessments. Though it was not explicitly stated, the implication was that these measures would either be newly developed or adaptations of existing measures. The performance tasks to be developed were to have the following characteristics: (1) a successful solution should require some sophisticated forms of mathematics; (2) the tasks should be understandable by students possessing a range of mathematical knowledge and of language abilities; (3) instructions should be open-ended enough that students would come up with innovative strategies; (4) students should be asked to show their work in sufficient detail (with rough drafts, etc.) that someone could follow the justifications for the answers; (5) tasks would be translated into the students' native languages and be presented in a range of media (e.g., paper and pencil, video tape, computer animations); (6) students should be encouraged to show their solutions in either language, through a similar range of media; and (7) if a group of students work together on a task, they should describe their relative contributions to the final product.

The portfolio assessments to be done were described in similar detail. The content areas to be addressed would be numbers and number sense, discrete mathematics, geometry and measurement, probability and statistics, rational numbers (including decimals and percents), algebraic reasoning, and other advanced forms of mathematics. The work would be scored on four dimensions: (1) mathematical content (forms of mathematics demonstrated); (2) mathematical communication (quality of communication); (3) conceptual knowledge of mathematics (evidence of understanding mathematical ideas, using algorithmic solutions, etc.); and (4) mathematical literacy (skills necessary for work, home, and citizenry). The tasks and scoring would be based on the judgments of highly skilled math teachers about what should be expected at specific ages.

The panelist who discussed science suggested that measurement should include both objective measures of content mastery and exploratory performance tasks. She pointed to two existing measures as models, but indicated that new measures would need to be developed. The panelist listed the National Assessment of Educational Progress (NAEP) science assessment and New York's fifth-grade Program Evaluation Test (PET) as models. She suggested that assessment instruments be prepared which could be used at "key benchmark years" (such as grades 4, 7 and 10), and that parallel measures be developed in "at least Spanish and Chinese." In addition, she suggested that attention be given to the relative difficulty of tasks for students of different backgrounds, and that provisions be considered for the modification of testing procedures for students with limited literacy skills.

The two other panelists discussed outcome measures in core content areas in a more generic fashion. One panelist suggested the use ("with possible modifications") of assessments in mathematics, language arts, and science being developed by the New Standards Project. These assessments include multiple choice, short answer, longer answer, and portfolio assessments for students at grades 4, 8, and 10. The objective questions for the math assessments are available in both English and Spanish. As an alternative, the same panelist suggested using the California Learning Assessment System, which has been abandoned by the state.

The other panelist described an approach that measures the cognitive demands of a task with a task structure and scoring rubric that can be implemented in different subject areas. This approach had been through validation, generalizability, and instructional sensitivity studies. The panelist suggested that most performance-based assessments had not been through extensive validation studies, and thus that some caution should be applied to their use. In her work, she indicated that the issue of language dependence had been attacked by using a range of approaches, including mini-glossaries, demonstrations, and visual materials. Students can also use different modes for responding.

c. Administration of Outcome Measures

The panelists identified a number of issues relating to the administration of LEP student outcome measures for assessment within core content areas. The most important of these relate to: (1) the use of sampling; (2) the language of administration; and (3) how to deal with LEP student absences.

The panelists emphasized that the question of who is assessed should be based on the purpose of the assessment. As one panelist stated it, "For national-level program evaluation and research purposes, it would be acceptable to sample students, perhaps testing LEP students only at key benchmark years.... For program evaluation at the local level, more data collection points might be recommended." The same panelist suggested that annual assessments in key grades could be given outside the usual intensive testing period in the spring.

Two other panelists suggested that matrix sampling approaches might be applied, in which different students are given different assessment tasks. In this way, a broader universe of content areas could be covered in the assessment. Depending upon the purpose of the assessment and the subgroups about whom conclusions are to be drawn, matrix sampling might be very difficult, however.

Panelists suggested a number of approaches for dealing with the issue of language in assessment. For certain of the existing measures, Spanish versions are available. For new assessment instruments, panelists suggested that Spanish and perhaps other language versions be developed, though as one panelist suggested, "Special attention would have to be given to exploring the relative difficulty of specific tasks for students of different cultural and educational backgrounds."

Panelists also proposed a range of approaches for adapting administrations for students with limited language proficiency. Among those approaches were: (1) simplifying the language of English versions by using active voice, present tense, and short sentences; (2) presenting tasks through a range of media; (3) allowing students to use a range of media in their responses; (4) using a screening procedure (English proficiency test) for determining when assessments in English should be used; (5) allowing students to select the language of administration; (6) providing an oral reading of the assessment content; and (7) allowing students to use dictionaries.

In order to deal with LEP student absences, panelists suggested that the scheduling of make-up tests is extremely important. They also suggested that the use of portfolio assessments can ameliorate the problem, because students do not need to be present for an assessment session. Their portfolios can even be assessed after they leave school.

d. Analysis of Outcome Data

Panelists identified three types of issues relating to how academic achievement outcome data from LEP students should be analyzed and interpreted: (1) the research models which should be used in analyzing achievement data; (2) the standards which should be used in judging LEP student outcomes; and (3) how academic growth among LEP students should be determined and judged.

Panelists suggested a range of approaches for analyzing LEP student achievement data. Included among these were the traditional pre-post comparison group design in which the comparison group is either other LEP students not receiving special services or non-LEP students in the same age-grade cohort. One panelist suggested a planned variation study in which different types of services to LEP students are compared on a pre- and post basis. The panelist recognized, however, that planned variations of services often vary on more dimensions than were included in the design. A number of panelists proposed designs in which the background characteristics of LEP students (length of time in the U.S., language abilities, etc.) and/or measures of opportunity to learn are "controlled for," either through separation into groups (i.e., blocking) or through statistical controls such as analysis of covariance or multiple regression. There was agreement among panelists, however, that large-scale longitudinal studies of LEP students are impractical because of their mobility and the difficulty in defining patterns of services across a long period.

A number of panelists suggested that LEP student achievement in core subjects should be compared with national standards. Among the standards cited were the NAEP proficiency levels, the performance standards built into the New Standards Project, and curriculum standards being defined in specific subject areas. Those proposing the use of standards, however, suggested that LEP student backgrounds and opportunities to learn need to be taken into account in judging the effectiveness of programs in helping students reach standards.

Academic growth among LEP students in core content areas was defined in a number of ways by panelists. It was defined in terms of: (1) increasing mastery of challenging content

as measured on criterion-based assessments; (2) increasingly similar levels of mastery in comparison with non-LEP students in the same school; and (3) changes in performance on academic achievement measures defined in terms of effect sizes. Panelists disagreed on the value of using effect sizes and other similar measures of program effectiveness. The strengths of such measures are that they provide a common yardstick for defining effectiveness. Their weaknesses are that they ignore the importance of small but significant changes, and that it is difficult to generate large effect sizes within the one-year time periods which are most practical to use for comparison.

2. Language Proficiency

The second category of outcome variables which panelists were asked to address involved language proficiency. In describing this category, the question provided to panelists listed both English and native language abilities, and also both oral proficiency and literacy in the language.

a. Key Outcome Variables to be Studied

In terms of language proficiency outcomes, three of the four panelists specifically included both English language proficiency and native language proficiency variables as possible student outcomes to be assessed. There was general agreement that language assessment included all four skills: reading, writing, listening and speaking.

The specific types of language outcomes to be defined as the focus of assessment were described as skills related to both academic and real-world tasks. For example, one panelist offered the following as some examples of the types of skills to be required and assessed:

> **Speaking:** express viewpoints effectively, communicate intentions and understandings, pose questions for clarification, understand communication rules for effective participation in group discussion; offer interpretations, clarifications; contribute new ideas in discussions.
>
> **Listening:** grasp concepts presented orally, understand clarifications when presented, attend and respond to the contributions of others in discussion.
>
> **Reading:** search for information, interrelate ideas, generalize, summarize, explain information.
>
> **Writing:** organize thoughts to express a point of view, or write a well-developed story, provide evidence for an argument or point of view, or interpret/explain information to others.

b. Measurement of Outcome Variables

The use of multiple measures and the inclusion of performance assessments were consistent themes within the panelists' responses. Two of the panelists specifically included

standardized tests of language proficiency in conjunction with performance-based assessments. The other two panelists focused on performance assessments.

Standardized tests. The standardized tests that were mentioned were the LAS and the LAB, for which there are both English and Spanish language versions. One of the panelists recommended the use of such tests together with performance-based assessments of academic language proficiency that are currently in the process of development.

The second panelist recommended use of the LAB as one means of measuring student gains in proficiency in the four language skills of listening, speaking, reading, writing. She described the LAB as useful in that it discriminates well at lower skill levels, which is good for assessing beginning ESL students. However, given that the focus in design of the LAB was on discriminating at low levels, the panelist noted that outcomes are difficult to interpret above the 40th percentile. In the Spanish version, the score distribution of the LAB is more normal, but there would be a question as to its appropriateness for Spanish-speaking populations in other geographic areas (the LAB was developed for use in New York).

With regard to writing assessment, the LAB is limited and does not ask for student writing samples. Noting this, the panelist refers to other models available for holistic scoring of writing samples. One example mentioned is the New York State writing tests (given in grades 5 and 8), and the Regents Competency tests given at the high school level. At grade 5, students select two writing tasks from among five categories (personal expression, personal narrative, description, process essay, and story starter); at grade 8, students select three tasks from among a different set of options. They draft and edit their work. The written samples are then evaluated using a holistic scoring rubric and are rated by multiple raters. At the high school level, writing samples are again obtained and scored, although at this level there are procedures for obtaining writing samples and scoring them within a number of languages other than English.

Performance-based assessment. The same two panelists who discussed the use of standardized assessments, also referred to the use of performance-based assessments. One mentioned the use of a series of writing samples over time being generated by a student as a means of demonstrating increasing English writing skills. The other referred to use of performance-based assessments of academic language use. Of the other two panelists, one focused on the use of performance-based assessments of native language literacy. This panelist explained his preference for omitting "traditional assessment of low-level content" by reference to three rationales:

(1) Low level content, such as algorithmic skills in mathematics and decoding skills in reading are not by themselves very important. This panelist sees these skills as best assessed in terms of the larger skills they support. For example, use of algorithms helps to solve problems; vocabulary is necessary to understanding a text.

(2) It is possible to understand how well a student can handle basic skills and knowledge by observing the quality of work within larger tasks.

(3) If realistic skills can be carried out without low level knowledge and skills, then the importance or relevance of those skills should be questioned.

This panelist proposed tasks to obtain an assessment of the level of literacy in their native language that would be at about the level expected for students at the same grade/age. Example tasks at the high school level would be a business memo, or a technical explanation of how to use a piece of equipment, etc. The tasks would need to be judged by expert teachers/informants.

The fourth panelist demonstrated a similar focus on performance-based assessment. She pointed to examples of tasks for assessing oral and literacy skills, in which students' work is judged at a variety of levels, including lower level skills such as decoding skills or the use of synonyms.

c. Administration of Outcome Measures

The appropriate assessment procedures are likely to vary based on the purpose of the assessment. Decisions regarding which students should be tested and the schedule for the assessment should be based on whether the assessment is for program evaluation or research purposes.

Assessment for program evaluation. One panelist noted that if the purpose of assessment is for local program evaluation, then assessing all students is most likely the best approach particularly since the assessment can also be used for exit purposes. The assessment schedule recommended by this panelist for language proficiency, including writing samples, was an annual one, with spring-to-spring assessment for continuing students and fall-to-spring assessment for newly entering students. However, she noted that under this approach, very mobile students would probably be under-represented.

The panelist also indicated that other assessments of language to examine age-appropriate oral communication and literacy skills, including those measured through demonstrations or performances, could be measured in key grades, and could be given mid-year rather than in the spring. She also suggested that for these assessments, a make-up testing period could be scheduled. If the purpose of assessment is for student evaluation as well as for program evaluation, then again it would be important to include all students and to test on an annual basis.

Assessment for research. This same panelist noted that for broader research purposes, it would be acceptable to sample students, perhaps only at key years (such as grades 4, 7, 10 or grades 5, 7, 9). Also, adaptations to adjust for differences in student background and ability levels were suggested. For example, the panelist recommended that, in the case of tasks assessing writing skills, non-literate recent entrants would be noted as present, but would not take the assessment. She suggested that adaptations could be made in the test questions for LEP students of different literacy levels. Also, key grades could be selected for the assessment as compared to an annual assessment schedule.

Concern for the quality of student sample work. One of the panelists emphasized the importance of ensuring that the quantity and quality of the student work to be assessed is given careful thought and selection. First, the work selected should sufficiently represent the major domains within the area to be rated. For example, in the case of native language literacy skills, it was recommended that at least one sample of student work be obtained for each of the different kinds of textual material that was determined to be important. Second, the samples of work selected individually need to represent important student work and should represent the best quality work that the student has been able to produce.

The implications of these requirements for the development of student samples is that teachers need to be trained so that they are able to help students select high quality work, can encourage students to produce their best quality work for the specific assessment tasks, can analyze and score the quality and characteristics of the tasks that are included in a student's portfolio of work, and can score (as opposed to grade) student work.

d. Analysis of Outcome Data

Most of the comments concerning analysis of language proficiency data were provided by one panelist. Where standardized language assessments are used, it was recommended that comparisons be made against the norming population. In the case of the LAB, comparison could be either with the English-proficient norms or with the LEP norms. Standards could be defined in terms of NCE gains, and the expected gains would vary by grade level (i.e., gain of 10 NCEs in K-4, 7 in grades 5-8, and 5 at the high school level). These differences in the standards were recommended in order to reflect the varying learning rates of students at different grade levels. Another recommendation would be to have the standards differ for students who enter with different levels of initial first and second language (English) proficiency, but identification of these students would be problematic, and it would be difficult to implement.

To examine student increase in mastery of English writing through the use of a series of writing samples, the panelist suggested that comparisons be made with a New York or NAEP-type proficiency scale, and that over time and across grades, student proficiency levels should improve. The standards for this type of measure would be proficiency ranges defined for different grade levels, and ideally should be defined as higher than minimum competency levels. For example, there may be different proficiency objectives set for students in grades 1-4, 5-8, and 9-12. The panelist also suggested that the standards to be applied could be set differently for students at different ESL levels, since beginning students will have had less time to learn English and its written conventions than students with greater English proficiency.

3. Behavioral Outcomes Related To Student Effort

A third category of LEP student outcome variables which panelists were asked to address concerned behavior indicating academic effort or motivation. The examples which were provided to panelists were attendance, engagement in class, and school dropout.

a. Key Outcome Variables To Be Studied

Panelists provided an extensive list of behavioral variables related to student effort and motivation. This list included:

- school attendance;
- homework submission rates;
- ratings of cooperation with other students;
- volunteering to take on additional academic assignments;
- volunteering to help other students;
- teacher ratings of effort devoted to studies;
- persistence on academic tasks;
- level of engagement in class;
- participation in school-related activities;
- enrollment in advanced classes;
- school dropout; and
- eligibility for further education.

b. Measurement of Outcome Variables

Only one of the four panelists chose to discuss in detail the measurement, data collection and analysis of variables related to student effort and motivation. He focused on student engagement and persistence in academic coursework.

In order to examine whether LEP students were taking challenging courses, he suggested gathering course enrollment information and disaggregating it based on student gender, social class, ethnicity, and language proficiency.

The panelist suggested measuring student persistence and engagement through student self-assessments of the efforts which they made in their academic classes. Concerning a course in general, he suggested asking students:

> "...how much they are encouraged to "think hard, dig deeply into a problem, stay with it," and whether they are encouraged to (and if they actually DO) contribute to the development of shared understandings in content."

Concerning specific samples of work, he would ask:

> "...how engaged she had been in the production of this product, how deeply she had gone into understanding its details, what ideas she thought she learned or used in doing the task, how engaging the task was, and whether this really represents her best work or if she quit when she thought it "good enough" (i.e., if she persisted with the intellectual content of the task)."

c. Administration of Outcome Measures

In defining data collection methods for his measures of student persistence and engagement, the panelist who addressed this topic proposed two models. He proposed that data on course-taking by LEP and other students could be collected from school records once each school year. For his measures of student persistence and engagement in the classroom, he suggested that student interviews be conducted at the same times as data are being collected about student performance (i.e., when performance tasks are being collected and scored or when portfolios are being evaluated).

d. Analysis of Outcome Data

The panelist who addressed this topic suggested some guidelines for analysis. For data on student course-taking, he proposed that course enrollments should be compared with overall school enrollments. In his words:

> "A rule of thumb would be that a school's diversity should be reflected in each of its courses, within a random margin of error. This would enable a school to track, at some gross level, how opportunity to learn is distributed among its students."

For data from student interviews on persistence and engagement, the panelist did not provide specific recommendations. In his general comments on the analysis of student data, however, he did stress three themes: (1) the evaluation of student outcome data should be relevant to a school's objectives; (2) serious efforts at school reform often start slowly and then build; and (3) a balance of academic, language, behavioral, and socio-psychological outcomes is to be preferred to a rigid focus on one category of outcome.

4. Psychological Outcomes

A fourth category of LEP student outcomes that panelists were asked to address concerned psychological variables. The examples which were provided to panelists included self-esteem, positive attitudes towards school, plans for future education, and cultural pride.

a. Key Outcome Variables To Be Studied

Panelists proposed a very wide range of psychological variables to be included as LEP student outcomes. Among them were:

- positive self-esteem;
- positive attitudes towards school;
- academic self-confidence and feelings of competence in school settings;
- positive attitudes towards doing school work well;
- intrinsic motivation to learn (not for monetary rewards);
- self-regulation including planning and checking;
- attributions for failure based on effort and circumstances;
- academic aspirations and plans;

- positive attitudes towards future personal and familial educational attainment and outcomes;
- a sense of responsibility and citizenship;
- positive attitudes toward classroom peers from diverse social, cultural, and linguistic backgrounds;
- positive attitudes regarding connections between school life and life at home and in the community;
- positive feelings about bilingualism and the use of their native language; and
- positive feelings about their country of identification.

One panelist described the importance of combining these factors by describing an "oppositional identity" to schooling which can occur in minority cultural groups. In such cases, students feel that they must sacrifice their cultural identity in order to succeed in school and in later life.

b. Measurement of Outcome Variables

Two of the panelists provided some discussion of how psychological variables might be measured. For measuring positive attitudes towards school and feelings of confidence and competence in taking on challenging school work, one panelist suggested the School Attitude Measure. In addition, she urged that a range of objective attitude measures be considered, and that one be validated or modified for use with LEP students. She suggested that different language versions would have to be developed, and that the English language version would have to use simplified language to ensure that most LEP students would be able to read them with the help of the teacher.

The other panelist suggested that survey questionnaires and interviews could be used to assess psychological variables. He suggested that survey instruments would have to be very carefully constructed because of different cultural values and understandings, and that they should be supplemented by interviews by well-trained people. He suggested that the interviewers should be aware of cultural issues and should be bilingual.

c. Administration of Outcome Measures

Only one of the panelists discussed the administration of outcome measures related to psychological variables. This panelist emphasized that because information would not be intended for making decisions about individual students, that it would be appropriate to sample LEP students where there were sufficient numbers in a school. She did believe that attitudes might vary across grades, however, and therefore suggested that samples at each grade be used, rather than sampling grade levels. She urged that a survey be completed once per school year, and that sufficient numbers of students be assessed so that control variables could be used in the analyses. In her view, teachers would probably need some training in sampling students and in administering the psychological measures.

d. Analysis of Outcome Data

Two panelists discussed the analysis of outcome data relating to psychological variables. One panelist proposed specific analytic approaches, while the other discussed how such analyses should fit within an overall study plan.

The panelist who proposed specific analyses urged that psychological outcomes be examined within the context of a number of other factors. She proposed that the analyses should control for time in the U.S., proficiency in the native language and English at program entry, parents' educational levels, and the student's grade level, both at entry and at the time of measurement. She also suggested that the analyses should explore the student's educational experiences outside of as well as in the U.S., including the years of schooling and the types of programs to which the student was exposed. She suggested that data across years be examined to determine if there are cohort effects.

The other panelist urged that psychological outcomes for LEP students be examined within the context of other outcomes. He did not believe that psychological outcomes should become ends in and of themselves, because he was concerned that students might have very positive academic self-concepts even in the presence of poor academic outcomes. He labelled this phenomenon "feeling good, doing bad," and thus proposed that academic self-concepts should not be addressed at the expense of other variables such as academic achievement and persistence.

5. Work Readiness Outcomes

The fifth category of LEP student outcomes which panelists were asked to address was work readiness outcomes. The examples which were provided were knowledge of career opportunities and positive job attitudes.

Panelists proposed the following variables within this category as appropriate LEP student outcomes:

- acquisition of basic skills;
- knowledge of career opportunities and their educational requirements;
- belief that everyday educational experiences prepare one for a career and work;
- knowledge of appropriate employment-related behaviors (workplace literacies);
- knowledge of how to apply for postsecondary education;
- plausibility of career goals;
- structured job experiences, such as internships or cooperative education;
- evidence of having performed community service;
- evidence of teamwork; and
- a sense of responsibility and citizenship.

None of the panelists provided additional detail on how to measure these variables or on how to collect or analyze the data.

C. The Relative Importance of LEP Student Outcomes for School Accountability

As a way of summarizing their comments about LEP student outcome variables, panelists were asked to select three to five specific outcome measures which they would include in an accountability formula for schools, and to divide a total of 100 points among those measures. Panelists were asked to do this twice, once for elementary schools, and once for high schools, and then to justify their choices and weighting. One of the panelists chose not to answer because he believed that schools should decide their own accountability priorities based on state policies and inputs from teachers, parents, and the community. Another panelist also questioned the assumptions of the question, but provided general reactions.

1. Elementary Schools

It is difficult to compare the responses of the panelists to this question because they used somewhat different categorizations in their accountability formulas. One panelist had seven components in an accountability formula for elementary schools, and divided points as follows:

Component		Points
(1)	increasing mastery of mathematics	20
(2)	increasing master of science	12.5
(3)	increasing mastery of social studies	12.5
(4)	increasing mastery of English	20
(5)	increasing mastery of native language	20
(6)	positive school attitudes/academic self-concept	10
(7)	good attendance	5

The second panelist used six components in the formula, and divided points as follows:

Component		Points
(1)	growth in performance in reading and language arts	20
(2)	growth in performance in mathematics, science, and technology	20
(3)	growth in performance in social studies	5
(4)	equal development of English and native language literacy	25
(5)	student engagement and persistence in school activities	15
(6)	socio-psychological and physical health and well being	15

The third panelist simply stated that she would weight English language proficiency and subject matter competence equally.

2. High Schools

As for elementary schools, panelists used somewhat different categorizations in describing their accountability formulas for high schools. One panelist used nine components in a formula, as follows:

Component		Points
(1)	mastery of mathematics	15
(2)	mastery of science	10
(3)	mastery of social studies	10
(4)	mastery of English	20
(5)	mastery of native language	10
(6)	thoughtful post-secondary plans	15
(7)	positive school attitudes/academic self-concept	10
(8)	good attendance	5
(9)	low dropout	5

The second panelist used five components and stressed the importance of each student's individual post-secondary plans. His formula was:

Component		Points
(1)	knowledge and skills needed to access post-secondary opportunities	50
(2)	broad-based literacies needed to participate in democratic and other social institutions	15
(3)	completion of high school	15
(4)	socio-psychological health and well-being	15
(5)	student rating of quality of high school experience	5

The third panelist emphasized English language proficiency and subject matter competence equally, but added graduation rates and eligibility for college as components at the high school level.

IV. CONCLUSIONS AND RECOMMENDATIONS

The purpose of this examination of LEP student outcomes was to provide recommendations to assist OBEMLA in providing guidance to researchers and evaluators. The comments provided by the panelists suggest that there are no simple answers to the questions which were posed.

In this chapter, Development Associates presents its conclusions and recommendations concerning LEP student outcomes. The chapter is composed of two major sections. In the first, we discuss a number of key issues relating to LEP student outcomes which were raised by panelists but which they did not discuss in detail. In the second section, we present our specific recommendations concerning LEP student outcomes.

A. Issues in Selecting Outcome Measures

In providing their responses, there were a number of important issues which panelists raised but did not discuss at length. Based on our analyses, however, we believe that these are key issues which must be confronted whenever a selection is being made of LEP student outcomes. We discuss four such issues:

1. What are the purposes of assessment?
2. Should native language development and the inclusion of examples using LEP student cultural backgrounds be goals in the instruction of LEP students?
3. What language(s) should be used for assessment?
4. Should measures of growth be used for LEP students instead of norms or criteria developed for mainstream students?

1. What are the purposes of assessment?

LEP students are assessed for a broad range of purposes. Most commonly, assessments are performed for student-specific purposes such as identification as a LEP student, placement in an appropriate program, or review of status for possible exit from special services. For this paper, however, OBEMLA requested that the focus be on outcome variables with evaluative and policy implications.

Even within this more limited area of focus, however, there are different purposes for assessment. A distinction can be drawn between assessments which are designed to measure the effectiveness of particular programs or activities (i.e., program evaluations) from those which are designed to determine overall levels of achievement at the school, district, state, or national levels (i.e., accountability assessments).

Perhaps the most important distinction between these types of assessment is in the desirability of program-specific versus generic measures of outcomes. For program evaluations, the ideal measures are those which most closely match the instructional goals

and activities of the project under study. The more generic are the outcome measures, the less likely are they to capture the unique accomplishments of a project. For accountability assessments, on the other hand, generic outcome measures are preferred because they allow for comparisons among educational units such as schools, districts, and states. Program-specific measures may provide some indication of success for an educational unit, but without comparable data from other units, it is difficult to put those results within context.

2. Should native language development and the inclusion of examples using LEP student cultural backgrounds be goals in the instruction of LEP students?

School policy-makers generally agree that the goals for LEP students should be similar to those for other students. There are two issues related to goals, however, on which there is no consensus: (1) Should students' knowledge of their native languages be maintained and expanded?; and (2) Should students' knowledge of their cultural backgrounds be reinforced and expanded?

These two questions lie at the heart of political debates concerning bilingual education. The answers to the questions do much to define the variables and specific measures which should be selected for assessing the results of school reform for LEP students. Because there is no consensus on these issues, selection of variables and measures need to be made on a case-by-case basis to reflect the particular practical and political realities of an assessment situation.

For example, though some of the panelists generally placed emphasis on the importance of outcome measures relating to native language proficiency, there are numerous situations and settings in which such an emphasis would not be appropriate. If schools are making no efforts to maintain and expand on knowledge of the native language, then the selection of native language proficiency measures as outcome variables for LEP students would be unfair and inappropriate.

Similarly, schools vary in the extent to which they value, reinforce, and expand on the cultural knowledge which LEP students bring into classrooms. Some schools "situate" their instruction of core content areas in the cultural backgrounds of their LEP students by using culture-specific examples. Other schools focus their instruction on the mainstream culture and do not use examples from minority cultures. The outcome variables and specific measures which should be used in these two types of schools should therefore also vary.

3. What language(s) should be used for assessment?

There is consensus among educators that an important goal of instruction of LEP students is the acquisition of English language skills which can be used in mainstream content area classes. What is less commonly agreed upon, however, is when and how English and/or the native language should be used in the assessment of knowledge and skills in core content areas. It is generally recognized that performance on academic assessments is a function of both content knowledge and the ability to demonstrate that knowledge using language. For most mainstream students, academic assessments in English primarily

measure content knowledge, while for LEP students those assessments measure both knowledge of English and content knowledge.

One of the panelists suggested that the fairest approach for assessing content knowledge is to perform the assessment in the language in which the student has the strongest skills. This straightforward suggestion raises as many issues as it solves, however. In many cases, assessment instruments are not available in students' native languages. Even when they are available in one or more native languages, equity concerns are raised concerning students from language groups in which assessments are not available.

It is also not clear when native language assessments should be used. If students are taught primarily in English and the assessments are in the native language, students may not be able to demonstrate the knowledge in the native language. Similarly, if the student is not literate in the native language, a written form of a test in that language is not appropriate. One panelist suggested that the only fair version of a test for such students might be an oral test in the native language.

In determining the appropriate language for assessment, both the students' language abilities and the languages used for instruction need to be examined. For most LEP students, however, who are in transition in their language skills, it is impossible to completely separate the effects of language proficiency and content knowledge in academic assessments.

4. Should measures of growth be used for LEP students instead of norms or criteria developed for mainstream students?

Educators of LEP students recognize both the advantages and problems associated with using mainstream norms and/or criterion levels in judging outcomes for LEP students. The use of mainstream norms and/or criterion levels emphasizes the responsibility of the school to provide challenging content and effective instruction to <u>all</u> students, especially those who are LEP. On the other hand, these educators recognize the special challenges which LEP students face in U.S. schools. A typical response has been to rhetorically apply the same standards for all students, but to make special adjustments for LEP students. This is done, for example, by excluding from comparisons students who are very limited in English proficiency, by examining outcome measures for LEP students separately from other students, and by studying change scores or growth curves rather than comparing with norms or criterion levels.

The use of change scores (the traditional approach), statistical controls for initial levels of performance, and growth curves (which assume at least three longitudinal measurements) all focus attention on changes in LEP student performance rather than on comparison with the performance of mainstream students. Panelists suggested that such attention to change is particularly important in examining the effects of specific interventions on LEP student performance (i.e., in program or project evaluations). If the focus of attention is on the overall status of LEP student performance, on the other hand, (i.e., as in a National Benchmark Study), the use of mainstream norms or criterion levels may be more appropriate.

B. Recommendations Concerning LEP Student Outcome Measures

We have organized our recommendations into two groupings based on the target audience for the recommendation. The first target audience is researchers and evaluators who are designing accountability measures including LEP students or who are examining the effects of specific programs or activities on LEP students. The second target audience is OBEMLA; we offer recommendations for what OBEMLA might do in its national leadership role to define and improve outcome measures for LEP students.

1. Recommendations to Researchers and Evaluators

In designing studies and evaluations including LEP students, researchers and evaluators should:

(a) **Consider the purposes of assessment.** If the assessment is for broad research or accountability purposes, measures should have wide-scale applicability and well-established reliability and validity. If the assessment is for the evaluation of a specific program or activity, the measures should relate as closely as possible to the content of the program or activity. (p. 8, 12)

(b) **Use a range of outcome measures.** In assessing the effects of school reform on LEP students, a range of measures should be used. The outcomes measured should include achievement in core academic areas, language proficiency in English (and the native language if possible), student behaviors related to effort and motivation, and psychological variables related to successful school achievement. For secondary-level LEP students, the outcome measures should also include variables related to readiness for work and for post-secondary educational programs. (p. 6, 10, 13, 15, 17)

(c) **Use multiple measures of outcomes if possible.** For example, in studying content area knowledge in core areas, the use of both standard objective measures and performance assessments will provide a more complete picture of student knowledge and skills. (p. 7, 10, 13)

(d) **Attempt to separate the effects of content area knowledge and language proficiency on measures of academic achievement.** This may be done by using native language versions of assessment measures, by using assessments and measures which are less language-dependent, by special administration procedures (oral testing, etc.), or by statistical techniques which control for levels of language proficiency. (p. 8, 13)

(e) **Decide on appropriate academic achievement standards for LEP students.** For program evaluations and for assessments involving LEP students with limited educational backgrounds, it may be preferable to use growth or change standards rather than standards or criterion scores developed for mainstream students. (p. 9, 13)

2. Recommendations to OBEMLA

In continuing its national leadership role in the education of LEP students, OBEMLA should:

(a) **Designate liaison persons to coordinate with the major standard-setting groups.** Two persons should be designated to work with standard-setting groups in the country (National Assessment for Educational Progress, New Standards Project, National Goals Panel, etc.). One of those persons should be an OBEMLA staff member and the other a researcher/educator who is knowledgeable about LEP students and assessment issues. These persons should serve as advocates for including LEP students in a fair way in assessments relating to national standards.

(b) **Assemble a task force to make recommendations concerning measures of language proficiency.** The task force should consider the strengths and weaknesses of existing measures of language proficiency. It should examine proficiency measures for English, and for languages other than English which are extensively used in the U.S. The focus should be on the usefulness of such measures for research and evaluation purposes. The task force should make recommendations concerning the development of new measures of language proficiency, including the nature and content of such measures and possible approaches which could be used for their development.

(c) **Identify and disseminate models for including LEP students in performance and portfolio assessments.** The models should include approaches involving students with different levels of language proficiency. These models should be disseminated within the Department of Education, to state education agencies, and to local school systems serving significant numbers of LEP students.

(d) **Develop and publish guidelines for the assessment of LEP student achievement.** These guidelines should be simple and realistic and should use as a starting point the guidelines developed by the Council of Chief State School Officers.* They should address when native language versions of tests should be used, when and if LEP students should be excluded from testing in English, and how data from LEP students should be interpreted and used. Such guidelines should be used in OBEMLA-funded research and in evaluations of OBEMLA-funded programs.

---

* Council of Chief State School Officers. (1992). *Recommendations for Improving the Assessment and Monitoring of Students with Limited English Proficiency.* Washington, DC.

APPENDICES

Appendix A: List of Focus Group Participants

Appendix B: Focus Group Questions

Appendix C: Written Recommendations from the Participants

Appendix A:

List of Focus Group Participants

Focus Group Participants

Dr. Eva Baker
Center for Study of Evaluation
10880 Wilshire Blvd., Suite 734
Los Angeles, CA 90024-1394

Dr. Richard P. Duran
University of California-Santa Barbara
Graduate School of Education
Santa Barbara, CA 93016-9490

Dr. Walter G. Secada
University of Wisconsin-Madison
Department of Curriculum and Instruction
225 North Mills Street
Madison, WI 53706

Dr. Judith Stern Torres
Torres and Associates
897 Schman Place
Baldwin, NY 11510

Appendix B:

Focus Group Questions

Focus Group Questions

1. In schools serving LEP students and which are undergoing school reform, what are the **most pertinent LEP student outcomes** that should be examined when considering the impact of such school reforms? (The government is interested in findings with evaluative, policy implications.) Please list and describe those outcomes within the following six content areas:

   (a) academic achievement in core subject areas;
   (b) language acquisition (English, native language, oral proficiency, literacy);
   (c) behavioral variables indicating student effort or motivation (attendance, engagement in class, dropout, etc.);
   (d) psychological variables (self-esteem, positive attitudes towards school, plans for future education, cultural pride, etc.);
   (e) readiness for the world of work (knowledge of career opportunities, positive job attitudes, etc.); and
   (f) other.

2. Please select **three or four specific LEP student outcomes** (one relating to academic achievement, one to language acquisition, and one or two others), and describe **how you would operationalize and measure the outcomes.** Describe in as much detail as possible the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures. For measures not involving language acquisition, please indicate how the measures deal with differences in English and native language proficiency.

3. For the same three or four LEP student outcomes, please describe **the appropriate assessment procedures and schedules for assessment.** This would include who should be assessed (are the measures appropriate for all grade levels, should any LEP students be excluded, should there be any sampling of students, classrooms, or grades levels), when and how often they should be assessed, and what special persons, resources, and/or staff training are required for the assessment. Given the high mobility of LEP students, what special approaches (make-up testing, etc.) should be used to ensure a complete picture of LEP student outcomes?

4. For the same three or four LEP student outcomes, please indicate **how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms.** What comparisons should be made (pre-post, comparison groups, national norms, criterion achievement), and what standards should be used for assessing effectiveness (how much of a change is needed to define effectiveness)? Should different comparisons and standards be used for different categories of LEP students? If so, how would they differ?

5(a). If you were to hold an **elementary school** accountable for their outcomes with LEP students, what **three to five specific outcome measures** would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.

5(b). If you were to hold an **high school** accountable for their outcomes with LEP students, what **three to five specific outcome measures** would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.

Appendix C:

Written Recommendations from the Participants

I.	Introductory Comments	C-1 - C-6
II.	Responses to Question 1	C-7 - C-19
III.	Responses to Question 2	C-20 - C-32
IV.	Responses to Question 3	C-33 - C-41
V.	Responses to Question 4	C-42 - C-50
VI.	Responses to Question 5A	C-51 - C-54
VII.	Responses to Question 5B	C-55 - C-58

SECTION I: INTRODUCTORY COMMENTS

Two of the focus group panelists began their responses to the questions by providing introductory comments, in which they presented issues which they indicated were important as preliminary considerations. Their comments are provided in this section.

Introductory Comments: Walter Secada

There are at least four considerations which should enter into any answer to these questions: a school's grade levels; whether it has a specialized student-oriented mission; whether the school is departmentalized or not; what the focus of the specific school reforms might be. I make this point because the answer to a question about "most pertinent LEP student outcomes" will vary contingent on a school's profile along the four dimensions of grade level, mission, departmentalization, and reform focus. While schools may benefit from the clear articulation of most-valued LEP-student outcomes; and while the government, other funding agencies, and other stake holders may also find value in clarity of purpose, one of the clearest and most consistent lessons which dates back to the Rand *Change* studies is that schools are organizations that actively adapt goals and programs. Hence, even while trying to articulate such clarity, we should recognize the factors which are likely to affect the relevance of these goals to the schools' own contexts.

Grade Levels

School-level student outcomes vary by grade. Traditionally, elementary schools have been for teaching initial literacy in reading, writing, social studies, and arithmetic--now, mathematics. In contrast, high schools are intended to develop more elaborated forms of academic knowledge in these and other domains. High schools have additional outcomes in the arts and physical education.

Successful student preparation for and transition into higher grades is a crucial outcome in elementary, middle, and junior high school. High school is to prepare students for and help them make a transition to work, postsecondary education, or the military.

As students get older and progress through the grades, schools are expected to help them learn about and cope with an ever-increasingly complex set of social issues. While elementary school students are taught about--and some would argue that these should be considered worthwhile outcomes in their own right--self-respect, avoiding strangers, the beginnings about sexuality, and how to avoid drug abuse; high schoolers have to deal with increasing violence and crime, drinking, the ready availability of many drugs and their abuse, social pressures for early sexual activity, sexually transmitted disease, and pregnancy.

School Mission

Many schools, especially magnet or other forms of specialty schools, have clear missions which vary in the amount of emphasis that they place on academic skills development. For example, Gary Wehlage found that some high schools that specialize in working with potential drop-outs will have more of a real-world skills or jobs orientation than a more comprehensive high school. Milbrey McLaughlin and Joan Talbert, in their studies of high school departments, found a high school which specialized in the arts. In this school, the mathematics department was very weak and it seems that the content of mathematics courses was a bit watered down.

Many schools are trying to develop and use cross-disciplinary curricula where content is integrated by solving real-world or realistic problems that draw on multiple disciplinary forms of knowledge. While these curricula are becoming more highly valued and are thought to be more authentic, it is also more difficult to pinpoint precisely the academic knowledge that they are intended to foster. Hence, student outcomes become more diffuse and more difficult to pin down, as in the outcome: students will solve real world problems. What this outcome means and how it becomes operationalized are subject to much debate and interpretation.

Departmental Status

Many high and junior-high schools, especially restructured schools, are modifying their departmental structures to provide students with more personalized experiences. Alternatively, middle schools, which often are organized as families or schools-within-school, are experimenting with content area departments. And in an effort to enhance the quality of subject area teaching, elementary schools are creating specialists' positions in reading, mathematics, and science. Regardless of these variations, there is a stereotype that non-departmentalized schools teach children, departmentalized schools teach the academic subjects. It would seem logical that educational objectives for LEP students would be adapted to these schools' dominant world views.

Reform Focus

Schools will vary in the focus of their reform efforts. Some schools might simply be trying to be more inclusive of LEP students while trying to maintain, or to change minimally, their current program. Other schools may be trying to change pedagogical focus (i.e., change some combination of curriculum, instruction, and assessment) while also trying to include LEP students. Indeed, [something] which I learned from the school-level study in the National Center for Research in Mathematical Sciences Education (NCRMSE) is that schools might change their mathematics curricula because the previous curriculum, in the words of one teacher, "simply did not work with our students." Curriculum change became part of an overall effort to make schooling more relevant to changing student populations. For

example, in addition to changing its curriculum, another mathematics department--whose minority student population was shifting from African American to recent Central American immigrants--was hiring certified mathematics teachers who were fluent in Spanish. In this school's case, the mathematics curriculum and the language of instruction were changing, even if teacher-dominated talk remained the norm during instruction.

Regardless the reasons why stake holders require something, schools and school personnel will interpret the goals for LEP students in terms of their dominant world views. Those world views will vary as a function of, at least, the above four characteristics of the schools. And in some cases, quite frankly, these recommendations will simply be wrong from the perspective of the school in question. That does not mean that the school is any less sincere in its desire to integrate LEP students into its mission. What it means is that school's context(s) make the use of these objectives suspect. Also, there are likely to be other school-level contextual features which would affect how school personnel interpret and adapt the following recommendations. Having made these caveats, I now proceed.

Introductory Comments: Judy Torres

Issues:

1) Instructional sensitivity: LEP students must be provided with instruction which is appropriate to or reflective of their level of cognitive and linguistic development.

2) Access: At whatever level of development, LEP students must have access to demanding content knowledge across the curriculum -- not only basic skills.

3) Appropriate measurement: students must be measured in the language which best allows them to demonstrate knowledge. Ideally, the instruments used for assessments of LEP students should be reflective of the instruction they actually receive, as well as the language of instruction (this may pose a variety of problems).

4) Standards: It is desirable that LEP students meet the same standards of academic excellence established for their monolingual peers across the nation. (see for example the NAEP performance standards). This was certainly the aim of the standard-setting efforts in New York City.

5) Equity: Grade-for-grade, it is desirable that LEP students master demanding content at the same rate as that of their schoolmates. This is clearly related to the issues of access and standards listed above. For this reason, I think that any responsible evaluation or research needs to examine "opportunity-to-learn" objectives (outcomes?) for each area to be assessed.

6) Cross validation: where possible and administratively reasonable, multiple measures should be sought.

(NOTE: I make these recommendations with full understanding of how difficult and costly they would be to achieve in practice.)

Some big and general questions: what is the actual content delivered in classrooms to LEP students? To what degree does it parallel that offered to non-LEPs? How much time do LEP and non-LEP students spend actively engaged in content learning? Through what modalities are students engaged in learning? To what degree are the needs of LEP students actively discussed and considered when schoolwide instructional decisions are planned and made? How and to what degree do staff serving LEP students interact and work with staff serving other students? How are potential conflicts over "turf" and resources negotiated? What and who supports the voices who speak for LEP students?

The biggest question: what resources would be available for the study to document these things?

Discussion:

Items 3 and 4 may create conflicting demands. Item 3 calls for assessments that reflect instruction which has been geared to the cognitive and linguistic needs of LEP students. Item 4 calls for instruction and assessments which make cognitive and linguistic demands of LEP students that parallel those for non-LEP students. These demands may conflict, given the varied educational backgrounds of LEP students and the resulting heterogeneity of students in bilingual and ESL classrooms.

In fact, mastery of demanding content is likely to take longer for those LEP students whose educational experiences are limited or very different from those of their classmates when they enter a U.S. school. The heavy emphasis on English acquisition also tends to take instructional time away from content-area instruction for LEP students.

This leads me to a number of conclusions:

1) that the performance "timetable" for LEP students who enter US schools with limited or different educational experiences may have to shift to reflect their need to acquire basic content. This should not be an excuse for teaching basic content to LEP students across the board -- only in clearly documentable cases. Appropriate documentation of student characteristics should be maintained in these cases.

2) that there be provisions for considerable flexibility in the types of measurements to be used to maximize the fit between measurement and instructional content.

3) OBEMLA, the Department of Education, or others interested in funding broad-scale research efforts will have to consider issues of generalizability in implementing conclusion 2.

4) To do this right, we should continue to look at former LEP students until they leave school (long-term outcomes). Administratively, however, this will be a nightmare. Perhaps we need to write separate groups of objectives for newly-entered LEP students/ students with some numbers of years in a US school / former LEPs).

Final considerations as to the **purposes** of assessment:

1) I am making the assumption that the lowest level of Federal concern is for the success of programs and services, rather than individual students.

2) I am assuming that there are still Federal concerns for policy issues, including questions about which program works best for whom and that English language acquisition is the major concern.

Additional challenges to making appropriate assessments:

1) even within a school or program, the instruction which a student actually receives may vary from year to year, or even within a class.

2) the educational history and preparedness of LEP students must be taken into account when measures of progress and mastery are made and interpreted.

3) the questions are complicated by issues of academic level. Are we talking about year-to-year progress? Or the longitudinal outcomes of a LEP student's entire academic career? At what point do we no longer include a former LEP student?

SECTION II: RESPONSES TO QUESTION 1

Response from: Eva Baker

1. ***In schools serving LEP students and which are undergoing school reform, what are the most pertinent LEP student outcomes that should be examined when considering the impact of such school reforms? (The government is interested in findings with evaluative, policy implications.) Please list and describe those outcomes within the following six content areas:***

a. ***Academic achievement in core subject areas***

Performance on standardized and validated performance measures that assess student's conceptual understanding, relevant prior knowledge, and ability to apply methods (and discourse) of the discipline. Uses of non-discourse based assessments may be given, but need to be validated. In the current reform, these assessments will be referenced to standards or curriculum goals, but these usually need further specification. These data should have at least some comparative component. Exclusion rates and reasons need to be provided.

b. ***Language acquisition (English, native language, oral proficiency, literacy)***

Outcomes in reading, in literature and content, writing, and oral language for academic and real world tasks.

c. ***Behavioral variables indicating student effort or motivation (attendance, engagement in class, dropout, etc.)***

Effort, persistence, attendance, dropout, eligibility for further education.

d. ***Psychological variables (self-esteem, positive attitudes towards school, plans for future education, cultural pride, etc.)***

Self-regulation (metacognition) including planning, checking, cognitive strategies, motivation and attribution (of failure).

e. ***Readiness for the world of work (knowledge of career opportunities, positive job attitudes, etc.)***

Plausibility of goals, evidence of teamwork, acquisition of basic skills, understanding of options and roles.

f. ***Other***

Instructional experience and exposure as a key explanatory variable.

Response from: Richard Duran

1. ***In schools serving LEP students and which are undergoing school reform, what are the most pertinent LEP student outcomes that should be examined when considering the impact of such school reforms? (The government is interested in findings with evaluative, policy implications.) Please list and describe those outcomes within the following six content areas:***

a. ***Academic achievement in core subject areas***

Viable possibilities include:

Language arts and mathematics standardized test scores in English or Spanish, if Spanish is the language of instruction.

Science standardized test scores in English where reform efforts include attention to science.
Performance based test scores in language arts, mathematics, and science (where emphasized) to the extent such tests are available and have proven accurate.

Portfolio assessments in language arts, mathematics, and science (if appropriate). Also, thematic project-level portfolio assessments may prove viable in school systems experimenting with thematic instruction as part of educational reform.

b. ***Language acquisition (English, native language, oral proficiency, literacy)***

Viable possibilities include:

Standardized English proficiency instruments such as the LAS or LAB emphasizing English oral proficiency.

Standardized Spanish oral proficiency assessments (LAS) for students instructed in Spanish.

Performance based assessments of academic language proficiency in English (or Spanish) being developed by the Council of Chief State School Officers (contact Lily Wong Fillmore, UC Berkeley).

c. ***Behavioral variables indicating student effort or motivation (attendance, engagement in class, dropout, etc.)***

attendance, homework submission rate, scores or ratings on cooperation with other students, volunteering to help other students or to take on additional academic assignments.

d. ***Psychological variables (self-esteem, positive attitudes towards school, plans for future education, cultural pride, etc.)***

positive attitudes towards the importance of doing school work well, positive attitudes towards classroom peers from diverse social, cultural, and linguistic backgrounds, positive attitudes regarding connections between school life and life at home and in the community, positive attitudes towards bilingualism and maintenance of a first language while acquiring a second language, positive attitude towards future personal and familial educational attainment and outcomes.

e. ***Readiness for the world of work (knowledge of career opportunities, positive job attitudes, etc.)***

knowledge about the connection of educational attainment to pursuit of career options, positive belief that one's everyday educational experiences prepare one for a career and work; knowledge about the relationship between educational attainment, occupational choice, and standard of living

f. ***Other***

ability to self-assess learning outcomes and academic/linguistic/literacy skill growth--this is related to academic achievement and language acquisition and pertains to acquisition of metacognitive abilities

Response from: Walter Secada

1. ***In schools serving LEP students and which are undergoing school reform, what are the most pertinent LEP student outcomes that should be examined when considering the impact of such school reforms? (The government is interested in findings with evaluative, policy implications.) Please list and describe those outcomes within the following six content areas:***

a. ***Academic achievement in core subject areas***

For elementary school, the core subject areas have traditionally been reading, writing, social studies, and mathematics (I purposefully did not say arithmetic). Of secondary importance in elementary schools, though in my opinion they merit being raised to core status, are science and technology. Physical education and the arts (drama, music, painting/drawing as expressive media) receive mixed importance depending on the socio-economic status of a school's students.

In middle school, the core subjects remain reading (though literature and grammar are beginning to replace reading as core), language arts, social studies (though these are usually thought of as being geography and history), mathematics, and science. Secondary are foreign languages and the arts. Health and physical education, while not of primary importance, are usually accorded greater importance than the arts and foreign languages.

In high school, the core subjects vary as a function of student tracks and student interests. For college-bound students, mathematics, English, social studies (history and geography), and the sciences are considered core subjects. For many students--both going onto and not going onto post-secondary education--the importance of physical education, especially extra-curricular sports, rivals the importance given to academic subjects. For students who are in vocational-technical tracks, while some importance is given to the academic subjects of mathematics, language arts, and social studies, at least as much is given to the technical areas which the person plans to enter.

Noting that there are (probably many) variations on what could count as core school-based knowledge, some consensus has developed that *all* students should leave school possessing what could be described as literacy in the broad areas of the physical and life sciences (mathematics, biology, physics, chemistry, physical science, and technology), the social sciences (civics, history, geography, and some psychology and sociology), the humanities (literature, drama) and fine arts. The desired levels of literacy can be thought of as what an individual would need in order to be a productive member of society who (i) could be an informed participant in this country's social, political, and democratic institutions (i.e., be an informed voter) and in its technically oriented work place (which could include the military), (ii) could

pursue goals for personal and professional growth, and (iii) could have reasonable access to later-life opportunities that are both personal and career oriented. By literacy I mean that an individual has some familiarity with a particular domain so that when he or she encounters a significant and realistic problem requiring knowledge in that domain, that person can make some sense of the problem, can use her or his knowledge of that domain to (a) generate new knowledge, or (b) figure out a way of solving that problem, (c) find someone else who can solve the problem, and (d) understand how the solution fits into the problem at hand. In my conception of literacy, detailed technical knowledge of a domain is not required; but understanding of some of the central ideas and how they are inter-related among one another and to realistic situations is required.

How these goals translate into specific learner-outcomes within each of the academic domains can be found in a raft of curriculum-reform documents: the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics*, the American Association for the Advancement of Science's *Science for All Americans*, the National Research Council's *Everybody Counts, Becoming a Nation of Readers, Civitas*, and other documents. NCTM argues that the fundamental student outcome is what they term "mathematical power," while AAAS has organized the sciences (including the social sciences) around the broad notion of literacy. In both cases, these ideas are similar to how I wrote about literacy, above. It should be noted, however, that many of the recently-written curriculum-reform documents suffer from excessive detail wherein the overall goals for education and their links to later-life opportunity are a bit blurred.

If these broad learner outcomes are desired for the general population, then it would seem that they should be no less desirable for the nation's limited English proficient students. The objectives for any school that is trying to include its LEP population, then, should be to foster the development of the knowledge and skills in these broad areas so that LEP students develop similar kinds of literacies as their English proficient peers.

Three points need to be made about these academic outcomes. While some people might argue that all students should have some literacy in all these areas, such a goal is likely to result in what currently happens in schools: broad and superficial coverage of lots of content has resulted in students not really knowing nor being able to apply very much. Current curriculum reform efforts are based on the coverage of less content, but coverage that has greater depth. On the other hand, for a student to focus on just one or two areas of study would result in an overly narrow specialization that would limit that student's ability to participate broadly in our society as is envisioned in many of the current reform documents. Hence, it would seem desirable to strike a balance between broad knowledge on the one hand and in-depth knowledge of 1 or 2 areas other the other. This, of course, is a variation of the argument for coverage of a core curriculum plus student choice to focus on those areas that are of interest.

Secondly, student literacy in these broad academic areas should be assessed in as direct a manner as possible. That is, assessment should try to place students in as realistic a setting as possible. Where breadth of knowledge is being assessed, students should be asked to perform realistic tasks whose solutions require the sorts of literacy discussed--for instance, to actually lobby for the passage of a bill in some governmental body. Where greater depth of knowledge is required, the tasks should be more complex and, in some cases, more reliant on detailed technical knowledge. Moreover, these assessments should focus on students' demonstrating what they actually know and can do; they should not be normative in the sense of ranking students along some continuum.

Thirdly, an excessive and over-rigid focus on the academic knowledge could lead to inequity. For instance, it is a simple matter to raise achievement scores or to raise the percentage of a school's students who perform at some *a priori* set level: simply get low achieving students to drop out of the school or not to participate in the assessments. For instance, when estimates of LEP-student performance are factored into NAEP results, California's placement relative to other states in reading drops. Schools in states that publish by-school test results argue that they should not be held accountable for their LEP and special needs student scores; hence they try to get those students' scores removed from the school's average or they encourage these students not to participate in the testing program.

b. Language acquisition (English, native language, oral proficiency, literacy)

Given the importance of communication in all of its forms for meaningful participation in the various spheres of our society, LEP students need to develop a broad base of language skills. While these skills are most likely to be developed in English, given the importance of English as the *lingua Franca* of this country, we need to remember that in a world of NAFTA, of a global economy, and of international tensions, the country's businesses and military also need people who are fluent and literate in other languages and who have a native-like understanding of the cultures and histories of other societies. Given that many LEP students enter school with command of a language other than English and also that many of these students have a working knowledge of at least one culture other than the dominant American culture--in the case of immigrant students, this knowledge of can be quite detailed and extensive--it would seem that schools should try to develop those students' knowledge and skills.

The reasons for developing of knowledge of and literacy in multiple languages and cultures for LEP students are similar to the reasons for developing other academic skills among accelerated students. First, many students enter school with initial intuitions and knowledge of the academic areas. Educational psychologists have argued that curriculum and instruction should draw on and develop those intuitions--the same reasoning should be true for knowledge of another language and culture. Furthermore, some students have out-of-school access to knowledge, technical skills, and tools that are highly valued; for instance, some students have computers, have

visited many places, or have learned a lot about some things due to curiosity or other advantages that they may have. Schools are always being encouraged to draw upon and develop that knowledge--the same reasoning should apply for developing an LEP student's knowledge of another language and culture.

c. ***Behavioral variables indicating student effort or motivation (attendance, engagement in class, dropout, etc.)***

Student effort and motivation are both a means to the end of academic learning and valued outcomes in their own right. Student effort and motivation are thought to result in enhanced student achievement. Moreover, motivation and effort result from schools' becoming more of a community as opposed to being rigid bureaucracies. The primary indicators of student effort and motivation include (a) student self-report, (b) teachers judgments, and (c) student behaviors: persistence in school and course taking, and participation in class and in other school-related activities. These latter behaviors are important since they provide additional evidence--beyond self- and teacher-reporting--that students are actually engaged in school.

Student persistence in school and course taking, and student participation in class and in other school-related activities are also indicators of opportunity to learn and the degree of student community in a school. Insofar as schools shift the focus of their assessments from being normative to being performance based, it will be difficult to track whether or not the so-called achievement gap is closing. Data about LEP student persistence and participation--insofar as they are also opportunity to learn data--provide an alternative means of monitoring whether a school is making progress in its efforts to include all students in its programs.

As noted earlier, an over-reliance on academic performance for monitoring student outcomes could result in structural inequities becoming part of schools. By documenting student persistence and participation at the same time as we look into student achievement, policy makers and other stake holders should be able to strike a balance.

While there is a danger in placing too strong an emphasis on academic achievement, there is also a danger in placing too strong an emphasis on student effort, motivation, affect, persistence, and participation. It is possible to sacrifice academic performance to these other objectives. For instance, in his study of programs for potential high-school drop outs, Gary Wehlage and his colleagues found out that the most successful programs were also those which had sacrificed any intellectual content in their students' schooling. The literature on secondary schools is full of examples where teachers and students strike a bargain: in return for student compliance with minimal norms of good behavior and participation, teachers reduce the intellectual content of the work that they ask their students to engage in. In *Selling Students Short*, Michael Sedlak and his colleagues documented the case of a teacher who actively tried to get some hard-working students to do less than the rest of the class since otherwise the other students would score less-well than their harder-working

colleagues. [What should] be sought is a balance: we want high academic performance of all students; on the other hand, we want all students to participate and persist, to be motivated and engaged.

4. *Psychological variables (self-esteem, positive attitudes towards school, plans for future education, cultural pride, etc.)*

Psychological variables such as self-esteem and positive attitudes are important as both outcomes and process variables. We do not want to sacrifice a student's sense of worth and self-esteem in the process of academic success. [...] Some, for example, Fordham [...] Ogbu, [...] Delgado-Gaitan, [...] Valenzuela, [...] Nieto, [...] Gibson, [...] Suarez-Orozco, [...] Matute-Bianchi, [...] Olsen, [...] Trueba, [...] Cummins, [...] Phelan, [...] Davidson, [...] Cazden, [...] Erickson, [...] Rosaldo, [...] Bartolome, [...] Trueba, [...] Rodriguez, [...] Romo, [...] Harris, [...] Carter, [...] Segura, [...] Zentella, [...] Vigil, [...] Foley, [...] Valdes, [...] Reyes, [...] Halcon, [...] Arce, [...] Ferdman, [...] Weis, [...] McCarthy, [...] Valadez, [...] Matute-Bianchi, [...] Ogbu, [...] Gibson, [...] Ogbu, [...] Trueba, [...] Delgado-Gaitan

needs to be gathered through carefully designed instruments, carefully critiqued for misunderstandings, and supplemented through the use of individual interviews by well-trained people. When such efforts involve LEP respondents, then it is important that the instruments and the person administering them--especially in the case of an interview--be bilingual.

As in the case for (a), (b), and (c), above, there is a danger in focusing on socio-psychological variables to the exclusion of other student objectives. That is, these objectives can become ends in and of themselves. As one commentator on the minority achievement gap between the United States and other countries noted, "research studies, testing programs, [and] ... [p]ossible misdirected emphasis on image and cultural awareness ... [to the] neglect [of] the basics" for post-secondary educational plans. But this should not happen if the impetus for attention to academic performance and persistence

9. Readiness for the world of work (knowledge of career opportunities, positive job attitudes, etc.)

More than anything else, the importance of readiness to work varies depending on a student's track and aspirations to take the opportunity. Since so many students are likely to be already employed while still in secondary school, it is likely that they will already have developed some practical ways to apply for a job, job attitudes, etc. Instead, given the high drop-out rate of LEP students, and the lack of evidence that vocational programs tend to focus more attention on retaining students or on developing LEP (or at-risk) students' literacy skills, it would seem that LEP students in vocational programs may be at greater risk than others for not developing the literacy skills necessary to apply for post-secondary education. Since so many LEP students are the first generation of people for whom college is even a possibility, it would seem that their families and other social support networks would not have the literacies necessary to fill out financial assistance and college application forms, the knowledge that not all colleges are equal, or how one chooses a college, and what one looks for when deciding where to apply and go to, and the knowledge of what courses are important for getting into college (for instance, that chemistry is more important than an additional class of some other subject

Response from: Judy Torres

F. In schools serving LEP students and which are undergoing school reform, what are the most pertinent LEP student outcomes that should be examined when considering the impact of such school reforms? (The government is interested in findings with evaluative, policy implications.) Please list and describe these outcomes within the following six content areas:

a. Academic achievement in core subject areas:

Outcome: Opportunities to learn

1) Students will have access to and participate in the full mathematics, social studies, geography, and science curricula. In other words, they will have access to challenging content -- the same content that should be made available to all students.

2) In schools with departmentalized instruction, LEP students will participate in classes with demanding academic content in proportions at least similar to, if not higher than, non-LEP students.

3) The language(s) and materials used for instruction in these subjects will be understandable to LEP students as soon as they are appropriately challenged.

4) The students' native language and culture will be positively reflected in instruction, classroom activities, and the school climate.

5) The needs of LEP students are consistently considered in school-wide academic planning and decision-making.

Desirable Student Outcomes (both individual and group):

a. LEP students will demonstrate increased number of challenging content core subject area content mathematics, science, and social studies).

b. LEP students should demonstrate parallel levels of content mastery [rate of performance or exposure] with [non-LEP students in similar circumstances, when the same basic instructional requirements are being met].

c. [Not true in increasing proportion of LEP students who are able to participate meaningfully in challenging instruction in the core areas through proficiency in the native home language. This may be best measured in the student's first language.]

6. As time in an English language school increases, differences in the performance of (former) LEP students and their monolingual peers on parallel assessments should diminish meaningfully.

b. *Language acquisition (English, native language, oral proficiency, literacy)*

Opportunity to Learn Outcomes:

ntly differentiated to meet a
ort the student through the
racy skills.

Possible Student Outcomes:

1. English acquisition: students will make significant gains in closing the gap between their performance and that of their monolingual peers in the four domains of English listening, speaking, reading and writing. (the standardized test view)

2. Over time, LEP students must develop an increasingly full range of age-appropriate English language and literacy skills. Again, these should include high-order competencies in listening, speaking, reading and writing English. (the performance view) A partial list might include:

Speaking: express viewpoints effectively, communicate intentions and understandings, pose questions to clarify information, understand and use communication rules for effective participation in group discussion, offer interpretations, clarifications; contribute new ideas to discussion.

Listening: grasp concepts presented orally, understand arifications when presented, attend and respond to the contributions of others in cussion.

Reading: use LEP texts, search for information, nterrelate ideas, generalize, summarize, explain information.

Writing: write to express thoughts, express nt of view, or write a well-developed idea, provide evidence for an argument, nt or point of view, or summarize information.

2. Native language proficiency: LEP students will demonstrate increasing speaking, listening, reading and writing proficiency in their first language.

guistic or academic support
eir non-LEP peers on tasks
Overall, the former LEP
iteracy as a final outcome

strate increasing speaking,
language.

Over time, LEP students must develop a full range of oral communication and literacy skills. For most students, these are probably most effectively developed and appropriately assessed in their first language.

Again, these must be developed in listening, speaking, reading, and writing. The tasks are similar to those posed for English proficiency. A sample might include:

Speaking: express viewpoints effectively, communicate intentions and understandings, pose questions for clarification, understand communication rules for effective participation in group discussion. Offer interpretations, clarifications; contribute new ideas in discussions.

Listening: grasp concepts presented orally, understand clarifications when presented, attend and respond to the contributions of others in discussion.

Reading: see NAEP tasks: search for information, interrelate ideas, generalize, summarize, explain information.

Writing: ability to organize thoughts to express a point of view, write a well-developed story, provide evidence for an argument or point of view, or interpret/explain information to others.

c. ***Behavioral variables indicating student effort or motivation (attendance, engagement in class, dropout, etc.)***

Important variables would include:

1. engagement with challenging content in class (once the opportunity-to-learn-variables had been measured and accounted for).

2. attendance: high attendance for both individuals and programs.

   a. dropouts: that the differential in four-year dropout rates between LEP and non-LEP students be eliminated.

   b. dropouts: that the overall four-year dropout rate be reduced to less than five percent.

3. post-high school program/school outcomes: LEP students should be engaged in meaningful activity: either in post-secondary education, full-time employment, job training which is linked to the real possibility of employment, military service, or some combination of these. (Also see below.)

d. ***Psychological variables (self-esteem, positive attitudes towards school, plans for future education, cultural pride, etc.)***

It would be useful to measure some of the following variables:

1. positive attitudes toward school
2. academic self-confidence and feelings of competence in school settings*
3. intrinsic motivation to learn, not for monetary rewards but because learning is engaging and important in itself*
4. academic plans and aspirations
5. a sense of responsibility -- for schoolwork, employment, and community. A sense of citizenship (I am working on some references for this.)
6. positive feelings about the use of their first language
7. positive feelings about their country of identification

* These are reported to be associated with positive academic outcomes in the literature on achievement among African-American students. This research might be relevant here, too.

e. ***Readiness for the world of work (knowledge of career opportunities, positive job attitudes, etc.)***

Some useful variables would be:

knowledge of career opportunities and their educational requirements; actual structured job experiences, such as internships or cooperative education; evidence of having performed community service; knowledge of appropriate employment-related behaviors; a sense of responsibility -- for schoolwork, employment, and community. A sense of citizenship. (see (d) above)

SECTION III: RESPONSES TO QUESTION 2

Response from: Eva Baker

2. ***Please select three or four specific LEP student outcomes (one relating to academic achievement, one to language acquisition, and one or two others), and describe how you would operationalize and measure the outcomes. Describe in as much detail as possible the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures. For measures not involving language acquisition, please indicate how the measures deal with differences in English and native language proficiency.***

I will describe areas in which think I have expertise.

In the area of content understanding and problem solving, we have developed an approach that measures the cognitive demands of a task with a task structure and scoring rubric that is implemented in subject matter as appropriate. For tasks of conceptual understanding, we prepare specifications of the outcomes desired, the list of source materials used in the assessment, and train raters in the implementation of the rubrics (which include the use of prior knowledge, resources, integration, misconceptions, and overall understanding). The rubrics have been developed from inferences of expert and novice production (rather than depending exclusively on teachers' judgments). These measures have undergone validation including studies of factor structure (do the rubrics work) studies of reliability (can the raters be trained to agree) revisions in the rubric to make them instructionally useful, generalizability studies, to assure that task sampling is equitable, and instructional sensitivity studies to be sure they will be sensitive to reform. Efforts have been made to attack the issue of language dependence in content assessment by looking at (and validating) approaches that use cognitive mapping instead of writing, in inserting mini-glossaries to help with difficult terms, to shift the resources students use from text based to demonstration and visual materials. In the mapping examples, which we have used in science and history, students rely on source materials and then either in paper or pencil or in technology (HyperCard) relate important concepts, events, and facts to one another using links that include specified relationships, e.g., LEADS TO. The mapping approach also is domain independent. It can be scored by computer. The benefit of the domain independent approach, especially for teachers in elementary school and of course for children, is that coherent models of learning are imp'emented across different subject matters.

A trade off is that the accountability measure will focus attention (and presumably teaching and learning) rather than try to represent the full range of instructional practice.

Most performance-based assessments have not been through extensive validation studies and may very well yield peculiar results. In order to deal with the issue of credibility and standardized administration, I propose the use of on-demand measures.

In the area of oral language, we have developed measures of children's ability to read and understand texts, involving individual test administration. Passages are selected at random from the beginning and ends of adopted texts and children are first asked to read them and then to answer short questions, beginning with literal and moving toward more inferential comprehension. Students work is judged in terms of their decoding ability,including their use of synonyms as well as the quality of their answers to questions. We also have writing assessment measures that are developed in much the same way as the content example above. Other areas in which we have learning involve the measurement of team skills,where students are given rather than need to construct messages to one another that show their team competence (flexibility, adaptability,encouragement, leadership, decisiveness). The provision of messages permits the use of computers to keep track of interactions and also reduces the language burden.

Response from: Richard Duran

2. ***Please select three or four specific LEP student outcomes (one relating to academic achievement, one to language acquisition, and one or two others), and describe how you would operationalize and measure the outcomes. Describe in as much detail as possible the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures. For measures not involving language acquisition, please indicate how the measures deal with differences in English and native language proficiency.***

LEP outcome areas selected for discussion: academic achievement in core subject matter areas, language acquisition, behavioral variables, and psychological variable (including other above).

Based on my familiarity with assessment design and education reform, I would recommend the use (with possible modifications) of assessments in mathematics, language arts, and science (and possibly applied learning) developed by the New Standards Project (contact either Phil Daro, University of California President's Office--Phil is Director of Assessment for the New Standards Project or Lauren Resnick LRDC, Univ. of Pittsburgh--Lauren is Co-director of the NSP).

These examinations are being developed in two systems. One assessment system is a reference exam in a given subject matter area allowing for the study of relationships between state-level and district level developed exams responsive to education reform and exams developed by the New Standards Project. An NSP mathematics reference examination has been developed and piloted involving a combination of multiple choice, short answer, and longer answer problems in mathematics at grades 4, 8, and 10. This examination is available in both English and Spanish. Development in English language arts and science (both exams in English) is underway and presumably reference exams in both these areas will be available in the future.

The second assessment system being developed by the New Standards Project is based on portfolio assessments at grades 4, 8, and 10. Portions of the portfolio system are relatively well developed.

Both on-demand reference examinations and portfolio assessments are tied to performance standards that include attention to lower level basic skills and conceptual understanding, problem solving, and conceptual application and communication in a subject matter area. (For your reference, I have attached a draft version of the NSP Standards as of 1/26/95--a new version is just out and I don't have this yet--it's similar to this one).

The strengths of the New Standards Project examinations are that they are based on student performance standards that are sensitive to curriculum standards advocated by teacher professional groups such as NCTM and NCTME. Public perceptions that this has been a strictly top-down process are not accurate. Development of the assessments and scoring rubrics has extensively involved teachers (many from the New Alliance for Schools representing roughly half of all school children in the U.S.). There is a relatively small but highly active group of bilingual and ESL teachers who have contributed to this effort, and in particular to the design of Spanish language assessments--contact Linda Carstens, San Diego City Schools or Harold Asturias, U.C. President's Office.).

As an alternative, the California Department of Education Office (contact Dale Carlson or Sue Bennet) have constructed assessment similar to those proposed by the New Standards Project as part of the disbanded California Learning Assessment System (CLAS). While California has abandoned the use of CLAS because of controversies arising about it's failure to quickly produce accurate student level scores and because of objections of religious reconstructionists, the exam system is inherently sound and could be adapted for use by others. Indeed, the NSP Project has negotiated access to the old CLAS language arts examination items as part of it's own effort to develop a language arts exam.

[In passing, note that the NSP Project is not a federal or state government project, though states and large school districts voluntarily contribute to it's revenues.] Administration and scoring of an NSP or CLAS based examination system would probably have to be modified to enable meeting the goals of OBEMLA studies--see next section for a discussion of some of these details. Most importantly, existing NSP and CLAS assessment would need to be supplemented by assessments and qualitative data collections and analyses sensitive to students' behavioral and psychological development. Actuarial methods should suffice for some measures--e.g. student attendance, but other assessments might be more complex--e.g. selected collection and analysis of videotapes of students' interaction. The latter should not be dismissed as part of small scale studies informing understanding of how students' self assess learning and how they learn to collaborate and cooperate with children. In my opinion well selected collections of classroom videos have great utility in communicating to parents and educators how children learn in a classroom. This cannot be conveyed as effectively by scores on tests or even complex products included in student portfolios.

In my opinion design of an assessment system sensitive to education reform implementations need to be responsive to the 3 issues raised in the standards movement: appropriate curriculum standards, performance assessment standards, and opportunity to learn standards. A key question is to conceptualize and evaluate how teacher staff development regarding curriculum standards and assessment standards relates to opportunity to learn. In attempting to design student assessments, attention should be given to assessing students' opportunity to learn and teacher staff development practices aimed at enhancing student opportunity to

learn. Incidentally, I am critical of "checklist" studies of opportunity to learn that just gather data on classroom materials and curriculum design. We need more in depth examination of issues using qualitative research on what actually happens in classrooms--at least on a case study basis.

Response from: Walter Secada

2. ***Please select three or four specific LEP student outcomes (one relating to academic achievement, one to language acquisition, and one or two others), and describe how you would operationalize and measure the outcomes. Describe in as much detail as possible the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures. For measures not involving language acquisition, please indicate how the measures deal with differences in English and native language proficiency.***

A. Academic achievement: Mathematics

I would assess LEP students' knowledge of mathematics by giving them a complex task whose successful solution required them to do some sophisticated forms of mathematics. The tasks should be understandable by students possessing a range of mathematical knowledge and of language abilities. The tasks should call upon students to demonstrate that they know and can do mathematics. Instructions should be open-ended enough that students would come up with innovative strategies. And students would be asked to show their work in sufficient detail that someone could follow the justifications for their answers. Tasks would be translated into the students' native languages and be presented in a range of media (including paper and pencil, video tape, computer assisted animations). Student should be encouraged to show their solutions in either language, through a similar range of media. If a group of students work on a very complex task, they should describe, at the end of their work, their relative contributions to the final product.

While I would expect students to turn in a highly refined and polished final product, I would also ask them to turn in rough drafts so that I could better understand their reasoning and thinking as they were creating the final products.

Secondly, I would ask students to submit portfolios of their work which demonstrate the kinds of mathematics that they know and can do in the areas of number and number sense, discrete mathematics, geometry and measurement, probability and statistics, rational numbers (including decimals and percents), algebraic reasoning, and other advanced forms of mathematics. I would then analyze the work which students submitted according to whether or not the tasks themselves require or support students' production of significant mathematics.

Finally, I would then score the actual student work along the following dimensions:

1. Mathematical content: what specific forms of mathematics are students demonstrating through this work? Is it content that is related to the forms of real world literacy outlined above?

2. Mathematical communication: what was the quality of mathematical communication that the student attempted?

3. Conceptual knowledge of mathematics: given the content which the student work draws upon, does the student provide evidence of knowing and understanding central mathematical ideas in that domain? Understanding can be shown through the elaborated nature of the students' responses; through relating important ideas to one another; and/or through the use of algorithmic solutions which could not have been completed without the person understanding how to implement that algorithm and the conditions under which the algorithm applies.

4. Mathematical literacy: does the student's work demonstrate that she or he has some of the mathematical literacy skills that are necessary for later-life, out of school participation in our larger society and its various spheres of work, home, and citizenry.

Note that, especially when samples of student work are collected, the student's work may simply not produce evidence to support many strong claims about what a student knows and can do. Though it may be the fault of the task, student work would still be scored low. The nature of what we ask students to do can support or it can constrain what a student can produce in response.

The kinds of tasks which would be administered to students, the analysis of collected tasks, and the scoring of student work would be pegged to highly-skilled mathematics teachers' professional judgments about (a) the kind of tasks that students of a given age should be expected to attempt and to complete and (b) the kind of work that students should be expected to produce. Also at the later grades, I would depend on the professional judgements of people from a range of walks of life to determine whether the mathematical tasks are similar to those which students might encounter later in life and whether those tasks require mathematical forms of literacy.

In my work with the Center on Organization and Restructuring of Schools (CORS), we have collected samples of student work, scored the tasks on how well they support student production of high quality work, and then scored the student work along lines similar to those outlined above. We have successfully analyzed and scored work by LEP students that was in English as well as work that was in Spanish. In the latter case, we simply translated the student work into English before having the teachers score it. Also, we successfully trained teachers to use general scoring rubrics for social studies and mathematics.

B. Language acquisition: Native language literacy

I would expect LEP students to be able to read, with understanding, a piece of text in their native language. For instance, I would ask a high school, Spanish speaking student to read an excerpt from something like the *Hundred Years of Solitude*, a

business memo, a technical explanation on how to use something (a dishwasher, a lawn mower, a VCR) or how to assemble something, and some other forms of writing. I would then ask the student to, in her or his own words, describe what was read. I would use a series of prompts to help the student provide an elaborated account for what she had read.

I would score tasks and student responses on whether they provided evidence of literacy that would be adequate for what one would expect of individuals at their respective grades and/or ages, as well as how well the individual understood the content of the materials which he or she had read. Once again, these are high inference judgments which should be made by expert teachers and informants with real world knowledge and understanding.

Note: these performance-based approaches to assessment, essentially omit the traditional assessment of low-level content. There are three reasons for doing so. First of all, the low level content--algorithmic skills in mathematics, decoding and vocabulary in reading--are valid parts of the school curriculum only insofar as they are part of something larger. That is, a mathematics algorithm is important only insofar as it helps on solve a real problem; vocabulary is important only insofar as it helps one to understand text (or play on a television game show). In my opinion, assessment should focus on what is important.

Second, we can get information about student knowledge and skills on low-level content from the actual work that students produce. That is, I can tell how well a student can do some basic computational algorithms if, when reviewing work on a task that required the production of an algorithm, by studying the quality of the algorithmic work in the context where it makes sense.

Thirdly, if realistic tasks can be accomplished without recourse to low level knowledge and skills, then the validity of their inclusion in the school curriculum is suspect. If at no time a student produces a paper-and-pencil algorithm because she has done all of her computations with a pocket calculator (and done them correctly), then the importance of these skills has been reduced. Likewise, if someone can write elegantly without recourse to esoteric vocabulary, than the importance of that vocabulary has been diminished.

C. **Behavioral variables: Student persistence and engagement**

I would assess how much students were engaged in and persisted in academic course work through a variety of mechanisms. First, I would gather course enrollment information and disaggregate it based on student gender, social class, ethnicity, and language proficiency. A rule of thumb would be that a school's diversity should be reflected in each of its courses, within a random margin of error. This would enable the school to track, at some gross level, how opportunity to learn is distributed among its students.

Second, I would ask students about their different courses. I would ask them about how much they are encouraged to "think hard," "dig deeply into a problem," "stay with it," and whether they are encouraged to (and if they actually DO) contribute to the development of shared understandings of the content. Third, with each sample of work that a student turns in for the assessment of her academic performance, I would ask a similar set of questions about how engaged she had been in the production of this product, how deeply she had gone into understanding its details, what (mathematical) ideas she thought she learned or used in doing the task, how engaging the task was, and whether this really represents her best work or if she quit when she thought it "good enough" (i.e., if she persisted with the intellectual content of the task).

D. **Adapting for variation in English and native language proficiency**

If I were assigning common tasks to students or using survey items to get at their socio-psychological well-being, I would simplify the language, use active, voice, present tense, short sentences, and make available bilingual versions to help ensure that LEP students understood the task requirements. Also, as noted above, I would allow for the presentation of the tasks through a range of media.

When training people to score student work, I would focus on keeping the various dimensions of that work as separate as possible. In the case of work that had been collected (i.e., tasks that I had not assigned), I would include a rating on whether the task in question was broad enough to allow for multiple points of entry, for students with varying levels of linguistic competence to engage in the task, and for people to express themselves in a variety of ways.

When scoring student work, I would recognize that the quality of the student's communication will be constrained by her or his proficiency in the language in which she communicates. Also, it is possible that an LEP student might not produce as elaborate a response as someone who is English proficient. Hence, I would encourage LEP students to use well-labeled diagrams for explaining work instead of extended narrative prose. Also, I would train the scorers of student work to keep distinct the four dimensions--communication, content, understanding, social literacy--that are meant to be captured by the scales and not to allow the quality of the students' writing to affect anything but their scores on the communication scales.

Response from: Judy Torres

2. ***Please select three or four specific LEP student outcomes (one relating to academic achievement, one to language acquisition, and one or two others), and describe how you would operationalize and measure the outcomes. Describe in as much detail as possible the measure(s) to be used for each outcome, how the measures would be (have been) developed, and what meanings and limitations of meanings are associated with the measures. For measures not involving language acquisition, please indicate how the measures deal with differences in English and native language proficiency.***

A. Outcome 1:

LEP students will demonstrate increased mastery of challenging content in science content (as an example). Drawing on NAEP as a model, these skills would include applications of basic scientific information and analysis of scientific procedures and data.

As their time in an English-language school increases, LEP students should demonstrate increasingly similar levels of content mastery when their performance is compared with that of non-LEP students in similar content areas within the same school.

(This might entail different time criteria for students with different entry dates and degrees of initial literacy.)

Operationalization/measurement. Drawing from NAEP and New York's 5th-grade Program Evaluation Test (PET) in Science, the ideal assessment would combine an objective measure of content mastery at various "benchmark" levels with exploratory performance tasks that would be administered and scored in the school.

I am thinking of this in terms of program evaluation and national-scale assessments, where we are sampling across a range of curricula.

How the measures were/would be developed. NAEP already exists, and a great deal of documentation already exists for it. Resources could be put to creating tasks of parallel content and difficulty for LEP students. Because LEP students tend to receive less exposure to science instruction than non-LEP students in US schools, the measures would have to offer a greater range of item difficulties for students of widely ranging levels of cognitive development.

The New York State PET test in science offers a model, but is available only at grade 5. It reflects the NYS science curriculum through grade 5, and is available in five languages. It also includes a series of standardized performance tasks which are

scored by classroom teachers. These are not excessively burdensome to administer and do not require extensive training for administrators. In addition, students have been reported to find them engaging.

Meanings and limitations. The limitations of any assessment designed for broad administration is that it is not likely to be equally applicable to students in all contexts or from differing educational backgrounds. Both the NAEP and PET science assessments clearly tap a range of knowledge, but probably reflect different curricular emphases. In addition, many LEP students do not have experience the same taught science curriculum as their monolingual peers due to an instructional emphasis on developing English proficiency. So the curriculum match may be problematic for some LEP students.

Why work for some common measure? I believe that many if not most schools and districts have not had the time, energies, or technical expertise to develop appropriate criterion-referenced measures for science. Rather than seek to synthesize an array of locally-developed measures, I would prefer to frame a range of relevant objectives around a core concern for challenging content measured across schools and contexts. This would permit comparisons of LEP students with their non-LEP peers in a range of contexts.

While the PET model is useful because of its combination of performance and objective tasks, it would require replication at other grade levels, perhaps 7 and 9. This would require resources (see below).

Language Variations. Parallel measures would have to be developed in at least Spanish and Chinese. Special attention would also have to be given to the exploring the relative difficulty of specific tasks for students of different cultural and educational backgrounds. Provisions should be considered for allowing modifications of testing procedures for students whose limited literacy skills may make it difficult for students to perform on the objective portion of the test.

B. **Outcome 2:**

LEP students will make significant gains in acquiring proficiency in the four domains of English listening, speaking, reading, and writing.

Operationalization/measurement. In good part for political reasons, I am opting for a standardized short-answer assessment as <u>one</u> measure of this objective. The SIAC's evaluation of language proficiency measures should inform this decision, but I have worked with the Language Assessment Battery and see it as a relatively reliable and short way to assess LEP students' growth in the domains of English proficiency. Geared to the objectives of New York City's ESL curriculum and the needs of its second-language learners, it contains a lot of "floor", which makes it good for beginning ESL students. As you know, it can be given in grades k-12. There are target norms for both English-proficient and LEP students.

Meanings and limitations. Its limitations are well known: its assessment of writing is limited; it does not ask the student to generate a writing sample. Because it was designed to discriminate at a low level, the outcomes become difficult to interpret above the 40th percentile. Use of the LAB (and the other proficiency measures) will require careful thought about how LEP students who "outgrow" them would be tracked from these measures to those designed for monolingual English speakers.

Because students reach the proficiency cut-point on the speaking subtest fairly rapidly, students reaching this point would not have to take this subtest again. Other types of speaking performances could be substituted.

Language Variations. Because of its considerable floor, the LAB can accommodate students with very little initial proficiency in English.

NOTE: I have not discussed the LAB Spanish version, but it makes parallel task demands, and has a more normal underlying score distribution. Although it is widely used, its appropriateness for Spanish-speaking students of other geographic areas would need to be considered.

C. **Outcome 3:**

Over time, LEP students must develop a full range of age-appropriate English language and literacy skills. These should also be measured through demonstrations or performances. One example is writing.

Students should demonstrate increasing mastery of English writing by generating a series of writing samples over time.

Operationalization/measurement. Models for a series of graded writing samples scored holistically already exist, again in New York State's writing tests (given in grades 5 and 8, and as the Regents Competency Test at the high school level). Both have rubrics for scoring a variety of writing tasks.

Meanings and limitations. The NAEP tasks include three types of writing tasks, informative, persuasive, and imaginative. The informative tasks include reporting and analytic tasks, either from personal experience or from given information. The persuasive tasks require the student to convince others of a point of view, or to refute an opposing view. Most papers are graded on levels ranging from "minimal" to "adequate"; fewer are "unsatisfactory" or "elaborated." Papers are not scored in the schools, adding to the cost of the assessment.

NYS grade 5 students select two writing tasks from five categories: personal expression, personal narrative, description, process essay, and story starter. Students draft and edit their responses. The essays are evaluated with a holistic scoring rubric. The use of multiple raters helps ensure score reliability. In grade 8, the test is similar, except that the students choose three tasks -- a business letter, a report

based on information given, and a composition based on one of four purposes (narration, description, explanation, or persuasion). Again, scoring procedures are similar.

Language Variations. At the high school level, New York has alternative procedures for writing samples to be administered and scored in a substantial number of languages other than English. Similar procedures should be developed for NAEP-type or other performance assessments of writing in LEP students' first languages.

D. **Outcome 4:**

LEP students will exhibit positive attitudes toward school and feelings of confidence and competence in taking on challenging school work.

Operationalization/measurement. Note: I would like to measure attitudes toward school, but am not familiar with any except the School Attitude Measure (and even this I have not used)... I think that academic self-concept and aspirations are probably very important to measure, but don't know much about measuring them in practice.

Resources permitting, it would make sense to consider available objective attitude measures and invest in validating or modifying one for use with LEP students of the major two or three language groups. This would not be inexpensive.

Meanings and limitations. While standardization is really important here, I am not sure if the constructs of school attitudes can be measured with the same items in LEP students (of different cultural backgrounds) as in their monolingual American peers. This would have to be investigated, as indicated above. The aim should be for parallel constructs, and not necessarily items.

Language Variations. Clearly, different language versions of would have to be developed. The English versions would have to use simplified language to ensure that most LEP students would be able to read them with the help of the teacher.

SECTION IV: RESPONSES TO QUESTION 3

Response from: Richard Duran

3. ***For the same three or four LEP student outcomes, please describe the appropriate assessment procedures and schedules for assessment. This would include who should be assessed (are the measures appropriate for all grade levels, should any LEP students be excluded, should there be any sampling of students, classrooms, or grades levels), when and how often they should be assessed, and what special persons, resources, and/or staff training are required for the assessment. Given the high mobility of LEP students, what special approaches (make-up testing, etc.) should be used to ensure a complete picture of LEP student outcomes?***

A key issue here is procedures for inclusion of all LEP and language minority children in assessments. A screening procedure for administering examinations in English will need to be developed. One criterion is to assess in English those subject matter areas taught in English. Another procedure would be to allow students to select the language of assessment. Another procedure would be to use results from an English proficiency test to inform the decision of whether to test in English. Yet another possibility would be to mediate administration of an English language examination by oral reading of the exam, allowing students to use dictionaries, and/or allowing students to respond in their language of choice (Contact Rebecca Kopriva, Delaware Dept. of Education regarding her studies of modified administrations of CLAS examinations in California).

Given that the aims of OBEMLA studies may be to evaluate programs, it would seem not desirable to require individual-level scores for every area of assessment. It will not be efficient or within resources e.g., to administer enough lengthy on-demand exercises to get accurate individual level scores. Matrix sampling of lengthy exercises, however, may lead to accurate scores at a grade level within a school or school district and allow comparisons across districts or schools.

Design of technically defensible performance assessments is difficult and should proceed only with input from experts in this area. Independent statistical consultants should review the design of assessments apart from experts who are already part of a research design team.

Response from: Walter Secada

3. ***For the same three or four LEP student outcomes, please describe the appropriate assessment procedures and schedules for assessment. This would include who should be assessed (are the measures appropriate for all grade levels, should any LEP students be excluded, should there be any sampling of students, classrooms, or grades levels), when and how often they should be assessed, and what special persons, resources, and/or staff training are required for the assessment. Given the high mobility of LEP students, what special approaches (make-up testing, etc.) should be used to ensure a complete picture of LEP student outcomes?***

A. Academic Achievement and Native Language Arts

In the case of academic performances, we would need sufficient samples of student work to be relatively confident that we had a fairly good representation of the major domains within that academic subject. For instance, the NCTM *Curriculum and Evaluation Standards* break up mathematics into about 9 broad, albeit different, domains. Hence, we would need at least one (preferably 2) sample of student work from each domain. In native language literacy, we would need at least one sample of student work for each of the different kinds of textual material that was judged to be important to different stake holders.

From the perspective of reliability, my understanding is that one should have at least between 5 and 8 samples of student work for scoring. For anything less than five samples, there could be some severe questions about the reliability of the scoring. Since we would need to gather at least 8 or 9 samples of student work to adequately students' mathematical performance, issues of reliability would be attenuated in this case. For other subjects, we would still need to gather enough samples of student work in order to reliably score their performances.

If samples of student work are being collected to create scorable portfolios, then it is important to ensure that the work is important and that it represents the best quality work that the student is capable of.

In both of these cases, teachers would need training on:

(a) how to help students select high quality work;

(b) how to encourage students to produce their best quality work for the assessment tasks;

(c) how to analyze and score the quality and characteristics of the tasks which are represented in the samples of student work;

(d) how to score student work. Use of analytic scales entails that teachers remember that they are *scoring* NOT *grading* student work. Also, teachers need to learn how to look for evidence in the student work of the sort that would be needed in order to ensure that the scales are applied meaningfully. Finally, teachers would need to learn that--unless the specific focus is language arts--the focus on LEP student work should draw a distinction between the student's language proficiency and his understanding of the task's mathematical requirements; hence, the separate scales.

B. Student Persistence and Engagement

Some of these data could be gathered once a year. It should be a relatively easy matter to track student enrollment and course taking. I have suggested that students be asked about engagement and persistence in academic work as part of the work that they produce for the academic and native language literacy tasks. Hence, the schedule for gathering those kinds of data would match the schedule for gathering data about student performance.

C. Sampling Issues

If assessment tasks are to be administered to all students, then the school should develop some of sampling procedure that would minimize the burden on any one student. The exact sampling would depend on the unit of analysis. For instance, if one wishes to make claims about the mathematical knowledge that is being developed in a given class, then it should be possible to spread out 20 tasks in such a way that each student has to work on 4 or 5 of them, and still be able to make statements about the class. On the other hand, if the school is the unit of analysis, then it should be possible to spread many more tasks, by grade level, across the school.

If the assessment strategy is to gather portfolios of student work, then I would recommend that the school gather all the relevant portfolios and use as much information as necessary to score the nature of the tasks that students are engaging in. Then, in order to minimize the intensity of the work for scorers, I would sample--either to the grade or to the class room--depending on the relevant unit of analysis.

D. LEP Student Attrition and Mobility

The response to issues of student mobility and attrition depends on the assessment strategy that is being followed. On the one hand, the school should be keeping data on student attendance, so that these data can become part of the data set that are used to evaluate student persistence and engagement. If performance assessments are being administered to all students, then there would seem to be nothing wrong with allowing students to do their tasks whenever they can. If portfolios of student work are being gathered--and their tasks and work are being scored--then the portfolios of students would are absent would have quite a few holes and the

portfolios of students who leave a school would stop at a particular point in the academic year. On the face of it, there is no reason why these students' portfolios should not be analyzed and scored the same as any other student's portfolio. The actual use and reporting of the resultant data, however, would depend on the precise questions that the evaluation was trying to answer.

Response from: Judy Torres

3. ***For the same three or four LEP student outcomes, please describe the appropriate assessment procedures and schedules for assessment. This would include who should be assessed (are the measures appropriate for all grade levels, should any LEP students be excluded, should there be any sampling of students, classrooms, or grades levels), when and how often they should be assessed, and what special persons, resources, and/or staff training are required for the assessment. Given the high mobility of LEP students, what special approaches (make-up testing, etc.) should be used to ensure a complete picture of LEP student outcomes?***

A. **Outcome 1:**

LEP students will demonstrate increased mastery of challenging content in science content (as an example). Drawing on NAEP as a model, these skills would include knowledge of basic information about the life and physical sciences, applications of basic scientific information, and analysis of scientific procedures and data.

As their time in an English-language school increases, LEP students should demonstrate increasingly similar levels of content mastery when their performance is compared with that of non-LEP students in similar content areas within the same school.

Appropriate assessment procedures/schedules: who is assessed? For national-level program evaluation and research purposes, it would be acceptable to sample students, perhaps testing LEP students only at key benchmark years (grades 4, 7, and 10?). For program evaluation at the local level, more data collection points might be recommended.

There would be a real benefit from coordinating the science assessment with NAEP, since this would facilitate comparisons with non-LEP students taking the NAEP science test. (I am conflicted about this, since many LEP students will not take much science in high school, and may not make it to grade 11.)

When? How often? Since the assessment would seek to measure general understandings and cumulative knowledge, annual assessments in key grades would provide meaningful data, and could be given outside the usual intensive testing period in the spring. For this and all the other measures proposed, a make-up testing period should be scheduled.

What resources would be needed? The first consideration would be whether and what modifications would be needed to use the NAEP or New York PET science tests as a model for this assessment. It is likely that either choice would require a commitment of resources for test modification and/or development, as well as

piloting and revisions in a sample of schools. This would require considerable resources which I cannot estimate at this point.

If such measures were administered, teachers would also need to receive training in administering the objective part of the test as well as setting up and scoring the performance part. In addition, test administrators would need basic materials and supplies for setting up the performance work stations. These primarily include simple materials such as thermometers and rulers. One major resource consideration would be whether non-LEP students would be included in the assessment. If the assessment were not NAEP or NAEP-like, this would increase the assessment costs.

What approaches should be used to ensure a complete picture of LEPs? This is a difficult question. The answer offered here will also be relevant to the other objectives below.

Both research and personal experience indicate that LEP students tend to be mobile. This may result in fragmented or discontinuous educational experiences. Some LEP students may have educational backgrounds which are very limited, or different from the curricula commonly taught in US schools. They also may experience a more limited curriculum than non-LEP students once they are in US schools.

Since it is unlikely that any study will be able to track large numbers of LEP students for long periods of time, it will be critical to collect detailed background information about these students for purposes of statistical control. This will have to include information on their educational experiences outside as well as in the US. This should include information about total years of schooling, and some indications of the type of curriculum or programs the student was exposed to. **Some measure of their language proficiency in their first language as well as English at the point they entered a US school will be essential**; information about their parents' educational backgrounds will be useful as well.

The best the study may be able to do is to sample cohorts of similar students, seeking to replicate findings for LEP students in key grade levels and contexts (in other words, do fourth-grade performance patterns hold over time? or do cohort effects predominate? In what program or SES contexts are particular outcomes observed?

In addition, the study will have to try to control for students' length of time in US schools, separating students by entry level (grades 1-4, 5-8, or 9-12). These variables will produce a large matrix, so either the total sample will have to be large, or the analysis will have to work with small numbers of strategic variables.

NOTE: it is unlikely that this information will allow the research/evaluation effort to say which treatment "works best." There is likely to be inconsistency in program treatments over time, making classifications of "who got what services" extremely difficult. What it will be able to do is indicate what LEP students know at key grade levels. It may also tell us, if there are resources to do parallel assessments among

non-LEP students, how the LEP students are faring in comparison to NAEP standards as well as to their non-LEP grademates.

B. **Outcome 2:**

LEP students will make significant gains in acquiring proficiency in the four domains of English listening, speaking, reading, and writing.

Appropriate assessment procedures/schedules: who is assessed? If the objective is local program evaluation, it is probably a good idea to assess all students, since the resulting information could be used for program exit purposes. If LAB were chosen, it could be administered to all LEP students in grades k-12. Individual subtests could be administered if meaningful measurement could not be obtained in a particular area. For example, non-literate students could take the speaking subtest. Given the political importance of this outcome, testing all students to generate gain scores could be helpful.

When? How often? Given that student mobility is an issue, and this objective calls for pre and post-test scores, I would test all LEP students on an annual basis, spring-to-spring for continuing students, and fall-to-spring for new entrants. LEP students should be retested on an annual basis for as long as they fall below the test's cut score and remain in the school or district. Once beyond the cut score, their progress should be followed with other measures.

Note, however, that such tracking puts a great burden on schools with limited student information systems, such as the ones I work with.

What resources would be needed? Testing materials in sufficient quantities and some training for classroom test administrators. The test administrator should be English proficient. Use of an existing measurement tool is relatively cost effective.

What approaches should be used to ensure a complete picture of LEPs? In this case, it is likely that the study will be able to generate some indicators of student progress in English acquisition over time. The test norms will also provide an indicator of student performance relative to the test's norm group. On the other hand, very mobile students will probably be under-represented. Also, since students' language acquisition is probably not linear, it will be important to control for grade level an the time of entry into a US school and their L1 and L2 proficiency at the time of entry.

C. **Outcome 3:**

Over time, LEP students must develop a full range of age-appropriate English oral communication and literacy skills. These should also be measured through demonstrations or performances. One example is writing.

Students should demonstrate increasing mastery of English writing by generating a series of writing samples over time.

Appropriate assessment procedures/schedules: who is assessed? For program evaluation and research purposes, it would be acceptable to sample students, perhaps testing LEP students only at key benchmark years (grades 4, 7, and 10? or 5, 7, and 9?). Clearly non-literate recent entrants would be noted for reporting purposes, but would not take the assessment. The test questions could be adapted for LEP students of other English literacy levels.

As in the case of science, there would be a real benefit from coordinating the writing assessment with NAEP, since this would facilitate comparisons with non-LEP students taking the NAEP writing test.

For purposes of program and student evaluation, as well as instructional planning, it would be extremely useful to assess all students' writing at least once a year.

When? How often? Since the assessment would seek to measure general understandings and cumulative knowledge, annual assessments in key grades would provide meaningful data, and could be given mid-year rather than in the spring. For this and all the other measures proposed, a make-up testing period should be scheduled.

What resources would be needed? Initially, a decision would have to be made whether, or how, to adapt an existing measure for use with LEP students. I would think that the nature of these assessments makes them readily adaptable and at moderate cost. The increased costs would lie in ensuring that teachers scoring these assessments were doing the scoring consistently.

If such writing assessments were administered, teachers would need to receive several training sessions in scoring the writing samples to ensure consistency in the use of the rating criteria. In addition, test administrators would need basic testing materials and supplies.

What approaches should be used to ensure a complete picture of LEPs? Sufficient numbers of students need to be surveyed to yield a sample sufficient for analysis. Survey responses will need to be analyzed in terms of the student characteristics described for Objective 1 above.

The analyses will have to control for student background variables: time in the US, L1 and L2 proficiency at entry, parents' education, and grade level, both at entry and at the time of measurement. They should also explore the effect of students' educational experiences outside as well as in the US, including information about total years of schooling, and some indications of the type of curriculum or programs the student was exposed to.

Cohort effects should also be explored in the data for successive years.

D. **Outcome 4:**

LEP students will exhibit positive attitudes toward school and feelings of confidence and competence in taking on challenging school work.

Appropriate assessment procedures/schedules: who is assessed? Since this information is not intended for making decisions about individual students, it would be appropriate to sample students where there are sufficient numbers in a school. Since attitudes might change across the grades, however, it might be preferable to sample students at each grade level rather select certain grade levels as "benchmark" grades. I would not recommend trying to do pre and post-testing, however, since I think that repeated administrations of the same questionnaire might invalidate the results.

When? How often? An annual administration of a survey in a sample of classrooms or with a sample of students.

What resources would be needed? Preferably, resources to adapt, pilot, and reproduce an existing instrument for use with LEP students, again preferably in the two or three most common non-English languages. Perhaps a publisher could be encouraged to take on this task.

In addition, teachers would need some training in administering the survey, and sufficient materials. They might need guidelines for sampling students.

What approaches should be used to ensure a complete picture of LEPs? Relatively large numbers of students need to be surveyed to yield a sample sufficient for analysis. The survey analyses will have to control for the student background variables previously discussed: time in the US, L1 and L2 proficiency at entry, parents' education, and grade level, both at entry and at the time of measurement. They should also explore the effect of students' educational experiences outside as well as in the US, including information about total years of schooling, and some indications of the type of curriculum or programs the student was exposed to.

Cohort effects should also be explored in the data for successive years.

SECTION V: RESPONSES TO QUESTION 4

Response from: Eva Baker

4. ***For the same three or four LEP student outcomes, please indicate how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms. What comparisons should be made (pre-post, comparison groups, national norms, criterion achievement), and what standards should be used for assessing effectiveness (how much of a change is needed to define effectiveness)? Should different comparisons and standards be used for different categories of LEP students? If so, how would they differ?***

I don't think this question is very sensible without understanding the particular context in which one is working Comparisons can be made among groups on a sampling, cohort basis. Pre-post comparisons, with complex assessments are pretty worthless, for the kids get frustrated easily. Comparisons among schools and between children of different program status need to be made for public credibility, but unless one simultaneously measures student engagement, opportunity to learn, effort, and other variables, inferences will likely be faulty. I assume that if we are operating in a standards referenced way, we will have some public standard setting (much improved over what we've seen to date) where levels of actual student work are valued and benchmarks set out. On the question of effectiveness, there are no good available models: arbitrary statements, i.e., 5% improvement, are not very sensible; effect size approaches ignore importance. The key issues are not how much "growth" do we want, but what patterns of improvement are acceptable. Not every kid nor every school should be expected to progress in the same way. We need multiple models to recognize improvement where it occurs.

Response from: Richard Duran

4. ***For the same three or four LEP student outcomes, please indicate how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms. What comparisons should be made (pre-post, comparison groups, national norms, criterion achievement), and what standards should be used for assessing effectiveness (how much of a change is needed to define effectiveness)? Should different comparisons and standards be used for different categories of LEP students? If so, how would they differ?***

As the literature review on institutional change indicates, adoption and implementation of reform is inherently processive. It will be critical that contractors have an explicit design that captures what schools and school districts are actually doing as opposed to what they say they are doing. As the recent NRC report "Assessing Evaluation Studies" states we need to have a clear understanding of what model of educational practice is actually being implemented at a school site--i.e. what is the treatment and how does this treatment represent differently coupled practices within a school and classroom and across different facets of community, home, and education policy institutions.

One key issue is whether OBEMLA studies will involve comparison groups. It would be desirable scientifically if schools or districts not systematically implementing reform could be compared to those which are not. A more realistic strategy would be to do a "planned variation" study where schools or district were compared according to design philosophies for reform implementation affecting language minority students. Within school district comparisons of school might also be pursued from a planned variation approach. Care should be taken to not presume that language service category is the only independent variable in explanatory modeling of program effects. For example, the nature and type of instructional activities students are exposed to may have a greater effect on outcomes than just language service program--and of course, there could be interactions between language service type and instructional activity type. (This is suggested by thinking about Warren and Rosebery's work on science learning). In this regard, one may advocate that all students regardless of language service should be expected to meet the same high performance standards eventually.

If a NSP based assessment system were used the results conceivably would be comparable to results obtained by states and school districts linking to this system (note "linking" is a technical topic and doesn't mean that a state or school district actually uses a full NSP assessment). An NSP-based system would only collect assessment information at certain grade levels (e.g., 4, 8, and 10) and this may be a limitation for OBEMLA purposes given the concentration of bilingual programs in the early elementary grades, but perhaps a hybrid assessment system might be developed administering Fall to Spring tests in grades 1-3 or in grades 5-6.

Setting standards for effectiveness of a reform program is a judgmental act. The evaluation literature suggests e.g., effect sizes of 3/4ths of a standard deviation or larger as indicators of noteworthy program effectiveness. Obtaining effect sizes of this magnitude within a school year is a difficult enterprise if standardized test scores are used. Cross year comparisons of outcomes at a given grade level might prove more sensitive despite the fact that student cohorts could be different. Analytic techniques such as Analysis of Covariance could prove useful, but expert statistical advice would be needed on the strengths and weaknesses of comparison procedures for outcomes in the same grades across years.

While large scale longitudinal studies may be unfeasible, small scale longitudinal case studies may prove very informative on how children's learning outcomes are affected as a reform initiative matures. Cross sectional studies won't be very sensitive for this purpose.

Response from: Walter Secada

4. ***For the same three or four LEP student outcomes, please indicate how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms. What comparisons should be made (pre-post, comparison groups, national norms, criterion achievement), and what standards should be used for assessing effectiveness (how much of a change is needed to define effectiveness)? Should different comparisons and standards be used for different categories of LEP students? If so, how would they differ?***

Typically, stake holders expect that, at a minimum, students are better off having been in a program than if the program had not been implemented. Hence, the minimum standard for any evaluation is based on the use of two comparison groups (those in and those not-in the program) with some sort of a pre-to-post design. While there have been alternative models--some of them quite complex--to this design, most seem predicated on this bottom line expectation.

Since I have seen so many evaluations that blindly follow a pre-to-post-with-comparison-group designs, I would like to stress that the use of student outcome data and the evaluation designs that would be relevant depend on each school's grade level, mission, departmental status, and reform focus, and on the interplay of the school's context with the funding-program purpose. For example, if the purpose of a program is to improve LEP student achievement, then a simple design comparing how students in the program fare relative to LEP students who were not in the program on some sort of pre-to-post assessment is adequate. However, the designs become more complex depending on the nature of the school's mission and how ambitious its efforts are. For instance, assume that a math-science specialty school is trying to **close** the achievement gap between its LEP and non-LEP students in mathematics and science. For this school, the relevant data would revolve around student performance in mathematics and science and not necessarily achievement data in any other subject! Note that the relevant comparison group would not be LEP students who were not enrolled in the program. Since the school's goals are to close the achievement gap and not simply to improve LEP student achievement, the relevant comparison group would be the school's non-LEP students.

A key step in the evaluation of any school's efforts would be the documentation of the school's fidelity to those efforts. For example, if a school's goals commingle the integration of LEP students into the school's mainstream classrooms with efforts at ambitious pedagogy (i.e., some combination of curriculum, teaching, and assessment) in those same classrooms, then the evaluation of the school's efforts would have to include a careful analysis of the quality of the pedagogy that students receive. On the other hand, if a school is committed to more traditional forms of pedagogy and simply wants to become more inclusive of its LEP students--for instance, all ninth graders in a given high school will enroll in algebra--then the evaluation of that

school's efforts should document how its students are being integrated. In these examples, I am following the most recent recommendations in the program evaluation literature which strongly suggests that evaluations are more useful to all stakeholders when they not only look at the program's "effects," but also help us to understand how those effects were arrived at.

I would also like to add a caveat: the more ambitious a school's efforts, the longer it will take to see any results and the more complex the evaluation. While there is evidence, for example, in projects such as *Cheche Konnen*, QUASAR (at the University of Pittsburgh's LRDC), Cognitively Guided Instruction (at the Wisconsin Center for Education Research, University of Wisconsin-Madison), Interactive Mathematics Project (at the Lawrence Hall of Science, UC-Berkeley), *Math and the Mind's Eye* and *Visual Math* (at the Mathematics Learning Center of Portland State University) and in many other places that schools can engage in ambitious pedagogy *and* include minority students in those efforts, these efforts have been possible with large influxes of money and with lots of time and support. The evidence concerning franchise operations (where an idea is quickly replicated at many sites) is much more mixed: schools--like any other organization--adopt the superficial trappings of an idea without truly understanding what it takes to turn that idea into a reality. Indeed, at the Center on Organization and Restructuring of Schools, we have tentatively concluded that there are three principles which undergird successful school efforts at authentic forms of pedagogy:

1. Commitment to an intellectual focus as opposed to a diffuse array of programs without much coherence;

2. Sustained and focused program development as opposed to short-lived efforts;

3. Communal and quasi-democratic forms of organization which are flexible in their accommodation to diversity among staff and students as opposed to rigid or overly hierarchical organizational structures.

Having made these observations, I would argue that the question of "how much change is needed to define effectiveness" does not make much sense if one takes seriously the fact that each school has its own unique contexts and will try to integrate LEP students in its own ways. It is only over the long term that we can retrospectively answer such a question for any single school. I have seen too many examples from the CORS and NCRMSE studies of school-level change and restructuring where schools start their efforts with a bang, only to then fall back. In contrast, I have seen schools that started their efforts very slowly and very deliberately, but have gained momentum.

I would recommend a different approach. I would ask schools to specify what they are trying to do in order to (a) reform their pedagogical practices and (b) provide meaningful opportunity to their LEP students. Then I would ask them about the information that would be helpful to them in gauging how they are progressing

towards achieving their goals. Thirdly, I would think about how to design an evaluation which provided the schools with the information they want, with information that would help to better understand their own process of change, and with information that would help them track student performance, behaviors, affect, etc. In other words, the evaluation system should provide meaningful feedback to the schools, it should be timely, and it should help them think about issues in ways that, maybe, they had not thought of before.

For federal and policy-related purposes, I would then try to find patterns across various schools to develop some sort of typology of efforts, information needs, designs, and relevant student outcomes.

Also, I would like to stress that the issue is one of balance. As noted earlier, too rigid a focus on a single category of outcome--academic, language, behavioral, or socio-psychological--can result in a school's going out of balance. Indeed, it is partly an overly narrow focus on academics over anything else that has led to charges that people who promote educational excellence are elitist; an overly narrow focus on English language development over all else has resulted in charges that promoters of English-only approaches to the education really do not care about LEP students' academic growth; and an overly narrow focus on access without regards to the academic quality has resulted in charges that people who are concerned about equity have contributed to the so-called decline on educational excellence. As John Goodlad pointed out in a *Place Called School*, we want it all. The issue for schools--and something which evaluations need to reflect--is to balance these competing demands based on their local contexts and situations. If policy makers and other stake holders fail to understand the need for balance and the tension that comes with trying to achieve a balance, the result could be an unmitigated disaster for a school's LEP students.

Response from: Judy Torres

4. ***For the same three or four LEP student outcomes, please indicate how the outcome information should be used for drawing evaluative conclusions about the effectiveness of school reforms. What comparisons should be made (pre-post, comparison groups, national norms, criterion achievement), and what standards should be used for assessing effectiveness (how much of a change is needed to define effectiveness)? Should different comparisons and standards be used for different categories of LEP students? If so, how would they differ?***

A. **Objective 1**

LEP students will demonstrate increased mastery of challenging content in science content (as an example). Drawing on NAEP as a model, these skills would include knowledge of basic information about the life and physical sciences, applications of basic scientific information, and analysis of scientific procedures and data.

As their time in a US school increases, LEP students should demonstrate increasingly similar levels of content mastery when their performance is compared with that of non-LEP students in similar content areas within the same school.

Comparisons to be made: In this case, the comparison would be with a NAEP-type criterion level to indicate content mastery.

In this case, comparisons would be made with the non-LEP grademates of the LEP students within the school building, controlling for time in a US school.

Standards to be used: Selected NAEP-type proficiency levels (from 250 to 350) could serve as standards for various grade levels or students with varying degrees of initial proficiency. (This objective as written does not call for gains in subject-area proficiency, although this could be done with pre- and post-testing.) Using this type of measure would require the use of proficiency ranges as standards for students at different grade levels.

The analysis could test the differences in proficiency/criterion levels for statistical significance, requiring successively smaller differences for students in the US for greater periods of time.

Different comparisons and standards for different categories of LEP students? Sliding proficiency criteria would make sense for this objective, since younger students should not be expected to attain the same levels of subject-area proficiency as older students. Here, it would make sense to set different proficiency objectives for students in grades 1-4, 5-8, and 9-12.

We have already indicated that it would be important to analyze the progress of LEP students in terms of their prior educational experiences, entry grade, time in the US, and initial language proficiency. If it were feasible, it would be important at least to develop sub-objectives for students with 1-3 years, 4-6 years, and 7 years or more in US schools.

B. **Objective 2**

LEP students will make significant gains in acquiring proficiency in the four domains of English listening, speaking, reading, and writing.

Comparisons to be made: In the case of the LAB, comparison would be with the norming population. Comparison could be made with the English-proficient norms, or with the LEP (target population) norms.

Standards to be used: The expected NCE gains would range from 10 in grades K-4, 7 in grades 5-8, and 5 at the high school level.

Different comparisons and standards for different categories of LEP students? As indicated above, standards should vary to reflect learning rates of students at different grade levels. They should probably differ for students of different initial L1 and L2 proficiency, but I am not sure how to identify these students or implement such a testing program in practical settings.

C. **Objective 3**

Students should demonstrate increasing mastery of English writing by generating a series of writing samples over time.

Comparisons to be made: Comparison would be with a New York- or NAEP-type proficiency scale indicating mastery. Over time and as grade level rises, student proficiency levels should improve.

Standards to be used: As discussed for Objective 1, using this type of measure would require the use of proficiency ranges as standards for students at different grade levels. New York State has minimum competency cut-scores and criteria, but I think that the criterion should be higher than minimum competency.

Different comparisons and standards for different categories of LEP students? Sliding proficiency criteria would make sense for this objective, since younger students should not be expected to attain the same levels of writing proficiency as older students. Here, it would make sense to set different proficiency objectives for students in grades 1-4, 5-8, and 9-12.

Similarly, the expectations should be different for students at different ESL levels, with the proficiency standard for beginning ESL students set lower than for students

at the advanced level. Regardless of grade level, beginning students will have had less time to learn English and its written conventions than students with greater L2 proficiency.

D. **Objective 4**

LEP students will exhibit positive attitudes toward school and feelings of confidence and competence in taking on challenging school work.

Comparisons to be made: One set of interesting and reasonable comparisons would be among various subgroups of LEP students, examining variations due to differences in current age, age-for-grade, current grade, gender, prior education, grade of entry, L1 and L2 proficiency at entry, types of programs received, parents' literacy, etc.

Another interesting comparison could be with non-LEP students at the same grade levels. This would be important, but might be difficult to implement without additional resources and incentives for participation by non-LEP students and their teachers.

Standards to be used: Standards could be developed along the following lines:

That students in the US for longer periods of time would indicate attitudes toward school which were at least as positive as those of students here for shorter periods of time.

That students who reported more participation in special bilingual or ESL programs should exhibit attitudes toward school which were more positive than students who did not report receiving such services. (This is unfortunately subject to error in recall and variations in what students think of as services, however.)

That LEP students would report attitudes toward school which were at least as positive overall as those of their non-LEP peers in the same school and grade.

Different comparisons and standards for different categories of LEP students? There are none for this objective that I can think of.

SECTION VI: RESPONSES TO QUESTION 5A

Response from: Eva Baker

5(a). ***If you were to hold an elementary school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

This is not a very good formulation, in my judgment. It isn't clear,for instance, if we are talking about multiple content measures. and whether it is anticipated that every child (or the school average) would need to hi teach equally. For reform for LEP, I would emphasize English language proficiency and subject matter competence equally. I would use other measures for explanation.

Response from: Richard Duran

5(a). ***If you were to hold an elementary school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

This question does not make good sense to me because it asks for weights regarding school accountability. I prefer not to answer it directly, because if I answered it directly it would distort my values significantly. Fundamentally, I believe that schools (either elementary or high school) need to decide which outcomes are accountability priorities for them given state policies, parents and teacher input and local community values.

Response from: **Walter Secada**

5(a). ***If you were to hold an elementary school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

The issue is one of finding a balance. However, for an elementary school, I would focus on (a) student **growth** in performance in the core areas of reading and language arts (20%), mathematics, science and technology (20%), social studies (5%); (b) the co-equal development of student English and native language literacy (25%); (c) student engagement and persistence in school activities (15%); (c) student socio-psychological and physical health and well being (15%).

In the CORS study of school restructuring, we found that elementary school teachers would not object to being held accountable for their students' individual **growth**. What they object to is being held accountable against absolute norms. More than one primary teacher said while she may want all of her students to be able to read at the end of the year, if a particular child enters her grade without knowing how to hold a book or the letters of the alphabet, then she has to back track to help that child catch up. At the end of the year, teachers would not object to being held accountable for how much that child had progressed; but, they argued, it is overly simplistic to hold them accountable for the child's continued failure to read on grade level. These teachers spoke about children's gazes following the flow of words, their being able to sight read specific words, children's realizing that reading is for the purpose of understanding text (not simply to decode), recognizing the letters of the alphabet, and knowing how a story flows as the kinds of things that this child would need to learn during that year. These teachers said that portfolios which documented how a child had grown over the year would be the best way of documenting their children's growth for accountability purposes.

Response from: Judy Torres

5(a). ***If you were to hold an elementary school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

The following constitutes at best a "wish list," since many schools do not have systematic assessments in place for some of these outcomes, and there are even fewer assessments in use across schools and contexts.

First: There should be some evidence that LEP students have access to challenging content.

Objective		Rationale
1. Measures of mastery of increasingly challenging content in the areas of:		
Mathematics:	20 points	a key foundation for logical and problem-solving skills, as well as a basis for science proficiency
Science:	12.5 points	essential content knowledge
Social Studies:	12.5 points	essential content knowledge
2. Measures of increasing mastery of advanced communication and literacy skills in:		
English:	20 points	essential for long-term success
Students' first language:	20 points	important base for developing L2 proficiency; build metalinguistic skills and self concept
3. Positive attitudes toward school/positive academic self concept:		
	10 points	could be critical for students' long-term success in school and career
4. Attendance:	5 points	necessary for learning to take place

SECTION VII: RESPONSES TO QUESTION 5B

Response from: **Eva Baker**

5(b). ***If you were to hold an high school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

Same deal. Plus eligibility for college, graduation rates.

Response from: Walter Secada

5(b). ***If you were to hold an high school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.***

For high schools, again the issue remains one of balance and of fidelity to the school's mission. This balance should incorporate the student outcomes and student choice. I would ask each student about her or his plans and aspirations (post-secondary education, work, military, and the like). Then I would assess whether the student had the knowledge and skills that are needed in order to access those opportunities (50%); note: I would include dual language competence among these outcomes. I would want to know if students had broad based literacies that are needed for meaningful participation in the our democratic and other social institutions (15%). I would hold a high school accountable for LEP student persistence and completion of high school (15%) and student socio-psychological health and well-being (15%). Finally, I would ask the students to rate the overall quality of their high school experiences as an independent accountability indicator (5%). In all of these cases, I would hold the high-school accountable for (a) maintaining common and high standards for student performance and (b) closing the gap between LEP and non-LEP students.

Response from: Judy Torres

5(b). If you were to hold an high school accountable for their outcomes with LEP students, what three to five specific outcome measures would you include in an accountability formula? How would you weight them? (How many points out of a total of 100 would you give each?) Please justify your choices and weighting.

First: There should be some evidence that LEP students have access to challenging content (i.e., percentages of LEP students participating in classes where advanced content is taught, and comparison of their participation and mastery rates with those of their non-LEP peers).

Note: I have included more than five here. It seems to me that there are many essential outcomes which need to be examined at the high school level. If you wish to truncate the list, select the outcomes with the highest point ratings: English language skills, mathematics, and well-articulated post-high school plans.

Objective		Rationale
1. Measures of mastery of challenging content in the areas of:		
Mathematics:	15 points	essential skills in logic and problem-solving; foundation for any other advanced work in math and the sciences
Science:	10 points	essential content knowledge
Social Studies:	10 points	essential content knowledge
2. Measures of mastery of advanced communication and literacy skills in:		
English:	20 points	essential for long-term success
Students' first language:	10 points	important as a personal and linguistic resource; builds metalinguistic understanding and self concept.

3. Students will have developed evidence of thoughtful post-secondary college or career plans:

	15 points	essential high school outcome; indicator of school counseling effort

4. Positive attitudes toward school/positive academic self concept:

	10 points	could be critical for students' long-term orientation to learning
5. High Attendance:	5 points	essential for learning; critical at the high school level
6. Low dropout rate:	5 points	measure of school efforts to keep students engaged and "on track"

yr3to17\sections

SIAC *Special Issues Analysis Center*

The Uses of Communications Technology for Language Proficiency Assessment and Academic Assessment

Task Order D190:
Written Focus Group Report

June, 1995

Development Associates, Inc.
Research, Evaluation, and Survey Services Division

This report was prepared for the U. S. Department of Education, Office of Bilingual Education and Minority Languages Affairs, under Contract No. T292001001, Task Order No. D190. The opinions, conclusions, and recommendations expressed herein do not necessarily reflect the position or policy of the Department of Education and no official endorsement by the Department of Education should be inferred.

Prepared by:

Annette M. Zehler

Development Associates, Inc.

Prepared for:

**Office of Bilingual Education and
Minority Languages Affairs
U. S. Department of Education**

Table of Contents

I. Introduction 1

II. Abstract 2

III. Findings 5

A. What Are Current Examples of Use of Technology for Assessment? 5

B. What Are Future Uses of Technology for Assessment? 9

C. What Factors Promote Use of Technology for Assessment? 12

D. What Factors Limit Use of Technology for Assessment? 15

E. Summary 17

IV. Conclusions and Recommendations 19

A. Implications of Use of Technology for Assessment of Limited English Proficient Students 19

B. Recommendations 20

Notes 24

Appendices

Appendix A: Focus Group Participants
Appendix B: List of Questions
Appendix C: Responses from the Participants

I. INTRODUCTION

A written focus group on the use of communications technology for language proficiency assessment and academic assessment was coordinated by the Special Issues Analysis Center for the Office of Bilingual Education and Minority Languages Affairs (OBEMLA) in the months of April through June, 1995. The purpose of the written focus group was to obtain recommendations regarding the potential of communications technology to address needs related to the assessment of limited English proficient (LEP) students. The recommendations developed out of the written focus group findings are intended to assist OBEMLA in providing national leadership in the area of education of language minority and limited English proficient students. The information obtained through this written focus group adds to the findings of a series of related research efforts which OBEMLA has defined over the past three years.[1] These efforts have focused on issues of accountability and assessment, especially within the context of institutional and instructional reform, and their implications for educators and researchers who work with limited English proficient students.

The written focus group involved four panelists who have been extensively involved in the use of technology and assessment, although with different emphases. One panelist has been primarily involved in the use of telecommunications for on-call interpretation services and language assessment. Two panelists were researchers whose work has examined closely the use of technology within education and its implications for instruction and assessment. The fourth panelist was also a researcher with recent work that focuses on the use of technology for portfolio assessment. The panelists were each asked to respond to five questions identified by OBEMLA regarding the current and future uses of communications technology for assessment. The purpose of this report is to provide an overview of the responses given and a summary of the recommendations made.

There are four chapters in the report. Chapter II provides an abstract of the responses to the questions identified by OBEMLA. The responses provided by the panelists are summarized in Chapter III, Findings. Chapter IV, Conclusions and Recommendations, presents recommendations developed from the responses of the panelists regarding the use of technology for assessment, especially for assessment of limited English proficient students.

There are three appendices: Appendix A provides a list of the panelists and their affiliations; Appendix B presents the questions as they were provided to the panelists; and Appendix C provides the panelists' responses to the questions.

II. ABSTRACT

The written focus group was organized around five questions which were submitted to the panelists. Shortened versions of the questions and a summary of the responses are presented below:

What has been your experience in the use of communications technologies in the assessment of language proficiency and/or assessment of academic skills?

Panelists provided descriptions of a range of current uses of technologies, including communications technologies. With regard to language proficiency assessment, the use of assessors located at a distance from those being assessed, and on-call via telephone or via telemedia was described. One example was the AT&T Language Line Service which identifies and screens interpreters through telephone interviews. Several examples of academic assessment were discussed, including computer-based portfolio development, such as the Digital Portfolio, and interactive videodisc technology. Also mentioned were the construction of a technology-rich environment that connects the home, school, and classroom, and includes telecommunications linkages via the Internet to resources outside the local area. The panelists emphasized that new methodologies of assessment are needed in order to describe and track the learning that takes place in these new learning environments.

How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in the assessment of language proficiency and/or the assessment of academic skills?

The panelists referred to the need to develop our understanding of assessment within the context of the new technologies, whether for language proficiency or for academic skills. A first step in developing this understanding is having a clear sense of instructional goals and the types of skills students should be building. Technology should then be examined for its usefulness as a tool for evaluation of those skills.

Panelists suggested that those schools that are already at work on reform efforts will be most likely to find communications technology useful, and will be the most successful in integrating the use of technology into the instructional program. Within schools involved in reform there is emphasis on communication, collaborative work, and student-directed inquiry which communications technology can support in a variety of ways. The panelists gave a number of examples for how technology can support reform efforts. These included providing access to persons outside the local area as resources, involving students in realistic problem-solving situations, and use of technology as a means of organizing, collecting, and sharing students' work. In addition, technology was seen as a potential resource for those developing assessments to use in sharing information on standards, on scales for rating, and on exemplars of performances.

In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?

Panelists noted that the question of cost-effectiveness is complex, particularly given that first and foremost what is needed is to evaluate the impact on student learning. Several of the panelists suggested that the costs of technology can be considerable, depending on the specific technology, and may not be possible for many schools and districts, despite the fact that costs for technology are continually decreasing.

Nevertheless, panelists suggested that by expanding the resources available and by breaking the traditional barriers of time and place, there can be considerable cost-effectiveness achieved. For example, in language proficiency assessment the use of assessors on an as-needed basis via telephone or via telemedia connections both can provide resources where none were possible before, and can save costs otherwise incurred in keeping persons on staff or bringing specialists in.

What specific communications technologies hold the greatest potential or promise for improving assessment of language proficiency and/or assessment of academic skills? What do you think are the limits of these particular communications technologies for the purpose of educational assessment? Does the use of technology lead to new models of assessment? What are implications for language minority students?

Many of the communications technologies that were mentioned as currently in use were again referred to by the panelists, especially in terms of ways in which use of the technologies could be extended in the future. For example, the use of distance-based language assessors via telephone or other media for the purpose of educational assessment was seen as one extension of current experience. The development of interactive videodisc technology into an assessment model for either language proficiency or academic assessment was also identified as a direction for future work. Finally, the computer-based portfolio system was seen as a system that could be extended more broadly, beyond the school, and used in other assessment settings. For example, student computer-based portfolios could be submitted for review by colleges, by future teachers, among others.

In general, panelists saw communications technologies as broadening the resources available for assessment of language minority students in a variety of ways. These included making assessment of native language skills possible where it might not otherwise occur, providing a means of assessing student work directly (e.g., by making it possible for assessors to see the actual work of the student), and by showing growth in skills within the context of the student's own beginning point. Also, examination of samples of student work at different steps in the development of a work product allow a reviewer to view the process as well as the outcome and thus develop a far clearer picture of the student's skills.

Panelists referred to a number of factors that could limit the effective development and application of communications technologies for assessment. These included not only the need for general access to the technology, and equity in access, but also factors such as the

training required for staff and the time for them to work with the technology, the need for the technology to be effectively integrated within the instructional approach and curriculum, and the costs of equipment and facilities to support the use of technology.

Panelists commented that with technology new models of assessment are possible since technology makes it possible to rethink our existing views of testing and of standards. However, for technology to be used effectively in assessment, new methodologies need to be developed and, in fact, are developing. Once developed, these new methodologies will offer the potential for better understanding of a student's processes in learning and of the nature of an individual student's work. A more individualized look at student growth and accomplishments, and observation of the steps students take in learning, will benefit language minority students in that their accomplishments and their growth in skills can be more closely documented.

What do you think are the future possibilities and the potential difficulties in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies? What suggestions do you have for how to best take advantage of possibilities or to overcome potential difficulties?

Panelists emphasized that use of communications technology for assessment should fit within the broader accountability system that is in place within a school or district. In addition, they emphasized the importance of informing and training staff; this might occur even before the technology is actually placed in a school. Once the technology is in place, there must be ongoing assistance available both for managing the hardware and software and for providing staff development on how to most effectively integrate the use of the technology within instruction and assessment. A suggestion was also made that the use of technology can be enhanced by agreements district-wide or between K-12 and higher education schools on some basic standards in terms of systems or approaches.

III. FINDINGS

In this section of the report, we present the findings of the written focus group for the five main questions considered by the four panelists. The questions posed to the panelists were focused on the use of communications technologies in particular, that is, technologies that allow communication between persons who are in different locations, and who communicate across different time zones. Interaction with others through the Internet, through telephone and combined telephone and video connections and distance-learning approaches are examples. The panelists discussed these and other promising technology-based assessments.

Two themes were prominent in the panelists' responses to the research questions. First, the panelists consistently pointed out the important relationship between instruction and assessment. Their comments indicated that the potential of technology is conditioned by its "fit" within an instructional setting. Thus, many of the comments described the use of technology for instruction as well as technology for assessment, and recommended the use of technology within settings where reform of instruction was also underway. Several of the comments predicted that greater access to learning and to assessment resources could be helpful to students, and to language minority students in particular, through the use of technology.

Second, the panelists anticipated that technology would bring not only significant contributions but also important changes to research and practice on assessment. They noted that much work needs to be done to fully understand the implications of technology, the best uses of technology within instruction and assessment, and the methodologies to be used in assessment using technology.

The following five sections provide an overview of the panelists' responses related to current experience and future prospects for the use of technology for assessment:

A. What are current uses of technology for assessment?
B. What are future uses of technology for assessment?
C. What factors promote use of technology for assessment?
D. What factors limit use of technology for assessment?
E. Summary

A. What are Current Examples of Use of Technology for Assessment?

Panelists discussed their experiences in the current uses of technology for assessment and instruction. They discussed a range of technologies, including but not limited to communications technologies. Their experiences have been varied, and the range of work described in the panelists' responses offers an important perspective on the possible uses of technology. The responses usually were focused on either language or academic assessment.

However, the technologies that were discussed have applications for assessment in both areas.

1. Uses of technology for academic assessment

Panelists saw considerable potential for the use of technology in academic assessment. One panelist pointed out as an initial perspective that different technologies make different types of assessments possible, and thus communications technologies and other newer technologies should add to the range of choices available in assessment. In making a similar point, a second panelist referred to a paper by Collins, Hawkins, and Frederiksen.* This paper presents the argument that different media (paper and pencil, computer, video) offer different views of students, and that, thus far, ideas of what types of skills should be observed as a measure of students' learning have been shaped by the assumption that the assessment would be carried out using paper an pencil. However, with technology, such as the computer and video, there are now additional possibilities being developed for looking at students' learning. For example, through the use of new technologies it is possible to assess a student's problem-solving within simulated environments, to assess ability to carry out a conversation within a realistic situation, or ability to work cooperatively with peers.

One panelist described experience in working with limited English proficient students as part of a Center for Children and Technology (CCT)** project. This project has involved a middle school with a student population that is 91 percent Latino, with 75 percent from homes in which a language other than English is spoken. The teachers and students were provided with access to use of computer technology both at home and at school and access to a variety of software. From their home and other locations, the students and teachers were able to access remote servers, on-line CD-ROM resources and encyclopedias, and could send e-mail locally and over the Internet. The students' involvement in the use of communications technologies at home and at school was correlated with improvements in academic outcomes, especially reading and writing skills.

Another panelist described work with a multimedia tool called the Digital Portfolio, which is used to record and organize student work. It is currently being implemented in six schools (one elementary, one middle, four high schools). One of the high schools is in an urban setting, serving a predominantly minority student population. The Digital Portfolio

---

* Collins, A., Hawkins, J., & Frederiksen, J.R. (1993). Three different views of students: The role of technology in assessing student performance. *The Journal of the Learning Sciences, 3*(2), 205-217.

** Education Development Center, Center for Children and Technology. (1994). *Union City Interactive Multimedia Education Trial.* Newark, NJ.

can store students' work in text, graphics, audio, and video form. Key features distinguishing the computer-based Digital Portfolio from a paper-based portfolio are:

- It provides examples of student work in a variety of media (i.e., text, video, audio, graphics);
- The work is organized on the basis of the "vision" that the school has for qualities/abilities a student should acquire; this vision "drives the assessment";
- The work can more easily be stored and viewed by a number of reviewers (e.g., parent, outside reader, college admissions officer, following year's teacher);
- As a portfolio, the samples of student work present a richer picture of a student's abilities than traditional assessment records (e.g., grades, test scores).

The Digital Portfolio is a software program designed for IBM and IBM-compatible machines with Windows. The structure of the portfolio begins with the definition of the goals of the school for students, i.e., the "vision" of the school. These become components of the "main menu", viewed at entry into the portfolio for review. Staff need to determine what types of tasks students can perform to demonstrate the skills or knowledge that fulfill the vision and then, once they have done this, to review the school's curriculum, scheduling, etc. to ensure that the schooling experience supports opportunities for students to carry out these types of tasks. For this reason, the panelist describes the Digital Portfolio as actually an overall strategy for school reform. When students select and enter examples of their work into their individual portfolios, they also need to select which goal and skills they believe that their work sample demonstrates. In this way, the Digital Portfolio is seen as a means of keeping the school's vision for reform "alive".

The panelists noted that a further advantage of technology use is that it can show the process as well as the product of student efforts, showing where the student started and the means by which the product was developed. These processes can be observed and can be included within assessment. For example, within a simulated environment, a student's steps to solving a problem or the number of times the student utilized any available help mechanisms all can be tracked. Or, within a Digital Portfolio, different stages in the development of a product can be shown. This aspect of technology use in particular may be of importance for limited English proficient students. The opportunity to include the processes as well as final products in a variety of media also provides the teacher with a more comprehensive picture of the students' skills. Thus, limited English proficient students' growth in skills can be seen by directly showing the different types of skills they have mastered over time. Through these type of assessments, the teacher can become very familiar with the students' skills and become better able to determine what the next steps should be for instruction. Also, access to a variety of media increases opportunities for limited English proficient students to demonstrate what they have learned.

Another panelist suggested that interactive videodisc technology is very promising for academic assessment, since persons are placed into realistic contexts in which they have

difficult problems to solve. These might be problems such as troubleshooting an electronic circuit, dealing with a difficult client, or manipulating objects to test understanding of Newton's law in different contexts. One example is a system called "Thinker Tools" that allows definition of activities for students to carry out, utilizing their understanding of Newton's laws.

2. Uses of technology for language assessment

Panelists observed that the advantage of the use of technology for language assessment is in the potential for placing students within realistic, novel situations in which they must use the language interactively. Effectiveness of language assessment would improve because the individuals would need to utilize their abilities in the language spontaneously; that is, they would not be able to prepare or predict. For this reason, one panelist felt that interactive videodisc technology would hold the most promise for assessment of language proficiency.

The panelist described work carried out at Northwestern University in which a system was designed to interact with persons visually. That is, it would present scenarios such as arriving at the airport, registering at the hotel, etc. in which persons would talk to the student and wait for a response. Although all of the inputs from the student were typed, the approach was one that tried to teach English by placing the student within realistic situations in which he or she needed to react. The system included videos of people talking to the student, so that there were opportunities to hear English spoken in these situations. There were various types of assistance available on-screen. For example, the student could see a typed version of what had been said in the video, and it was possible to get a translation of specific words used. The panelist suggested that this same technology and approach could be used for assessment, by scoring how well the student responds in the situations and how much help they need in responding.

Other systems mentioned were those that record a student's or trainee's voice and compares it to that of a model speaker, e.g., in terms of inflection or intensity. One purpose of this system was to train to a particular model of politeness for use in situations where the trainee would be needing to respond to persons asking for information or making complaints. Similar systems have been built (e.g., at Bolt, Beranek, and Newman) for deaf students to give them information on the placement of the tongue in pronouncing a specific sound. The system compared the tongue placement of the deaf student against a model so that the students could improve their production of speech sounds.

A second panelist described his work with a nationwide network of telecommunications-based interpreters (AT&T Language Line Services). This network uses communications technologies for distance-based testing and assessment to identify qualified interpreters for the Language Line services. In screening and assessing potential interviewers, a telephone-based Oral Proficiency Interview is conducted by a trained rater. With the applicant's permission, the interview is recorded and then scored a second time by a second rater. If both raters agree that the applicant meets the level required for interpreters, then the person is hired. Training is also carried out by telephone via a remotely managed conference call.

This involves as many as 39 students at one time who work with materials mailed out in advance. This type of screening and training has been ongoing for the past ten years. In this model, the distance-based assessment is especially appropriate to the type of work that the persons will be doing, i.e., it is all based on telephone communications. The assessment system has identified a pool of interpreters that cover the continent. Through use of telephone and computer, the system of interpreters can handle up to 140 languages, 24 hours a day, with an average interpreter connect time of about 45 seconds. This type of availability of language resources can have educational applications, and in fact has begun to be used by schools for interpretation services for parent-teacher meetings.[2]

A third type of assessment approach described was a multi-media computer-based portfolio system, such as the Digital Portfolio (described above). Although this was discussed with reference primarily to academic assessment, it is a model which easily could also be used for purposes of language assessment. The portfolio could show examples of written text, videos of conversations, presentations, or other demonstration of language skills. These presentations or demonstrations of language use can be available for review at different times by a number of reviewers. They can also be used for multiple reviews, utilizing different criteria for evaluation of the student's work.

B. What Are Future Uses of Technology for Assessment?

The panelists commented that the potential for use of technology is not always limited by the technology itself. There are a number of uses for which the technology exists, but the ability of schools and individuals to implement it does not exist. Thus, the potential for applying technology to assessment is qualified in terms of the potential for implementation. Teachers' knowledge and readiness to use technologies and the availability of physical facilities (e.g., a school building's capacity to accommodate needed wiring, etc.) are important preliminary supports for the use of technology.

Providing an overall perspective on the use of technology for assessment, one panelist pointed out that there are three kinds of roles that are possible for the use of technology. One role is to create environments or contexts in which language or other types of tasks are carried out. The context setting or the posing of questions for the learner could be done by a person at the end of a line, enabling the assessor to be in one place, and the person to be assessed in another. Or, as an alternative, these contexts can be presented in a computer environment, placing the person within simulated experiences or situations. These are more interactive than a paper and pencil assessment.

The second role for technology is in recording student performance. Video expands what is recorded far beyond that possible with paper and pencil. For example, gestures or facial expressions or intonation can be captured. The ability to work through a hands-on experiment can be recorded; or ability to interact, ask and respond to questions with another. This provides a whole different "window" on performance.

The third possibility is that technology might actually be used in analyzing or scoring student performance. The panelist noted that at present video or spoken English can only be scored by having judges rate the performance according to agreed-upon standards or predetermined criteria. However, within a computer task environment, it is possible to score how well tasks are carried out or to score how much help or hints are needed in order to accomplish the task. As the panelist noted, this would be related to the notion of dynamic testing discussed by Joe Campione and Ann Brown.

In describing future uses of technology, the panelists foresaw extensions of many of the currently available technologies. Again, the panelists made the point that educators need to focus on development of new methodologies for assessment in the context of the different views of student learning made possible by the new technologies.

1. Uses for academic assessment

One panelist commented that with technology our perspectives on assessment will change. For example, in working with computer-based portfolios, it is expected that the process of collecting the work for a portfolio will be as important a component of the assessment as the review of the portfolio. The process of identifying and selecting samples of student work can help a teacher to better understand the student's abilities, and to better view these in relation to the standards and goals of the school, that is, in relation to the school's "vision" for its students.

In this regard, future uses of technology can focus on showing the skills of the student in working through the process of developing a particular work sample. As an example, classroom activities can be recorded on videotape and multimedia to show how students began and what steps they took along the way. The particular pieces that will be important to include may be dependent upon the goal of the assessment, but with communication technologies and other technologies there will be a broader range of information to use in assessing. This would assist the teacher in becoming very familiar with a student and the student's abilities. In this way the teacher would be better able to evaluate the student's current level of understanding and to determine the next steps appropriate for that student's learning.

In looking to the future, the panelist predicted that the Digital Portfolio could take advantage of other technologies, such as the World Wide Web, which allows for transmission of text, video, audio, graphics between users. The use of networked technologies also can add to the possibilities of combinations of assessors and students sharing their ideas and reactions anywhere via the Internet. Thus a computer-based portfolio could be made very accessible to others. The student would be able to work in any medium while still having it available for review by individual assessors. These assessors will be able to look at the student's actual work, and make judgements based on the work itself, rather than solely on the judgments of others (e.g., college admissions/placement officers). However, the panelist cautioned that actually the most meaningful assessors are

those who are closest to the student (e.g., parent, teachers) and that putting a portfolio on a local area network would be more important than placing it in a form accessible by the Internet.

Another panelist also commented that through technology it is possible to see various versions of work as it is being created. These would be, for example, different drafts of a paper, or other records of student work, for a portfolio that a student is developing. Although not every draft would need to be seen and reviewed, the fact that there are versions of a work available emphasizes the fact that assessment is part of a "feedback loop", part of learning in progress. It thus makes possible observation of the processes that students use in developing the finished product or the steps that they take in solving a particular problem. As the panelist observed, this may be especially important for limited English proficient students since each student's abilities can be evaluated based on his or her own merits and point of progress. Also, by focusing on the student's work directly, assessment moves toward a focus on qualitative rather than simply quantitative summaries of a student's abilities.

Another panelist noted that communications technologies, particularly the Internet and World Wide Web, are able to support much more authentic learning practices. For example, via the Internet, students can engage in real-world science experiments, or take part in cultural exchanges. These are learning tasks which rely on interaction and communication with others and in which the type of learning that occurs cannot be adequately assessed by multiple choice standardized tests. As the panelist noted, such tests "do not do justice to the complexity of thinking and learning that take place in communications-based environments."

Similarly, another panelist predicted that through use of technology, assessment will no longer be viewed as generating sets of individual items; instead, assessment will consist of developing realistic task contexts or situations in which those being assessed must solve problems, answer questions or carry out commands. For example, by involving a language learner within specific situations, it will be possible to observe how that person can function interactively, and assess how well he or she is able to use language "on the fly", or in academic assessment, assess how well a student can apply scientific principles to actual problem-solving.

2. Uses for language assessment

One panelist pointed out that many "next generation" technologies are already available but are simply not being used. For example, it is possible for two persons to be communicating via telephone while simultaneously sharing word processing or data files on their computers. Even if the computers do not work with the same systems (e.g., one uses Word, the other uses Word Perfect) one person can export his/her file to the other person's computer and they can both look at and work with the same file. In this type of scenario, the two persons can work interactively, with one person typing in a sentence and asking the other to translate it, conjugate the verb in the sentence, or read the sentence aloud. This would be a one-on-one type of assessment situation. However, other types of communication options are also available and could be developed for use in education.

Service providers such as America Online have already made it possible for groups of persons (e.g., 20 or more) to converse via computer in a free form or moderated session that can involve persons from different time zones, or different geographic areas. These capabilities can be further explored for use in education and in assessment.

Also mentioned was the possibility of using a scholar in another country, e.g., Ukraine, to carry out assessment of a recently arrived immigrant via the use of PC-based telemedia. The panelist further commented that this type of linkage between the student being assessed and assessor has considerable cost saving implications. Through these types of linkages, a whole range of cost-intensive operations are avoided. For example, it is not necessary to try to locate someone locally who is qualified to do assessments, which is often extremely difficult to do, or arrange for a visit from an expert in the language who is located at a distance. Instead, this assessment model is, to use business terminology, an example of "just in time" assessment: the assessment resource is there when you need it through dialing the person on a video line, calling and then carrying out the assessment. The costs are the time-on-line required for the assessment. The assessment specialist can be located in one place and be on-call on a global basis. The panelist further pointed out: "with this approach, the number of speakers of a less-commonly-encountered language in a specific location becomes less of an issue because their needs can be addressed almost without regard to the 'clustering' traditionally required to achieve economies of scale."

The use of interactive videodisc and other technologies that involve learners within realistic situations was also seen as a means of obtaining a measure of a student's ability to use language. However, the means of assessing student language use in these situations still would rely on judges viewing video or viewing other samples of the student's performance in the language. The panelist noted it is necessary to develop further the approaches to examining and assessing student skills within these types of environments.

C. What Factors Promote Use of Technology for Assessment?

In their responses, the panelists noted a number of factors related to broader and more effective use of technology. Many of these factors center on the readiness of those who will work with the technology and the availability of support systems to help them use technology in the classroom. These factors are discussed below.

1. Ensure that teachers are receptive, informed about the technology, and trained in its use

An important factor in use of the technology is the teacher. As one panelist remarked:

> "the biggest risk (or downside) to the use of technologies lies in their uninformed use and the attendant unrealistic expectation that so often accompanies such use. For too many teachers in an educational setting, new technologies are like hand grenades

tossed into the classroom—you know something is going to happen when it goes off and it's never good...."

Thus, this panelist felt that providing proper introduction and orientation to new technologies is essential, and that this information should be provided before the technology is even through the door of the classroom. He also suggested that creative approaches to more comprehensively and quickly providing training to teachers should be developed. As one example, he suggested that a whole district commit to a specific technology and vendor, but then require that the vendor provide a long-term relationship in terms of training to every teacher and on-call accessibility for assistance with the technology.

A second panelist referred to a project in which homes, community, and school were all connected to one another by means of a computer network, so that it was possible to link up to resources from a variety of locations in the community. This greater access to technology resulted in increases in academic performance.* However, those implementing the project found that there was a very substantial need for ongoing assistance in a number of ways. Teachers needed training in how to integrate the use of computers within their instruction. There was a need for follow-up assistance in the use of the technology not only in terms of the instructional use but also in terms of the types of hardware and software used.

The panelists thus emphasized that for telecommunications-based assessments to be useful, it is important that those who will make judgements about the work be trained and supported in making those judgments. One panelist commented that although this requires extensive staff development, the benefits can be significant. In other comments, the panelist referred to the extensive staff training that has been carried out in support of reform efforts in Kentucky and Vermont, even though the use of technology was not included. Thus, the implication is that technology use requires training both for the new approaches and perspectives on assessment it involves and for the use of the new methods, equipment, and knowledge it requires.

2. Ensure that the use of technology is merged within the curriculum

One panelist commented that what are needed are "compelling reasons for students and teachers to use telecommunications technologies. A compelling reason is not, 'this will be useful later in life'. A compelling reason might be, 'this is something we can use for our current educational work'." If exploration and student-directed research are skills that are supported through everyday work in the classroom, then the use of technology to assist in that learning will be important. It will help students to do their work. Similarly, the use of technology to assess students' development of those skills will "make sense" since they

---

* Education Development Center, Center for Children and Technology. (1994). *Union City Interactive Multimedia Education Trial.* Newark, NJ.

will be assessed in precisely the types of higher order skills they are working to develop. Other panelists in general emphasized the need for the technology to provide assessment outcomes that make sense within the curriculum.

3. Begin with schools that are involved in reform efforts

Related to the above point, one panelist suggested that in designing new forms of assessment it would be useful to begin by working with schools that are already thinking about reform, and designing new forms of assessment. These would be the schools for which the technology helps teachers and students to do what they are already trying to do in the classroom. Schools already involved in reform are more likely to be carrying out activities designed to foster greater communication, exploration, and interaction with new information. They have educational goals for which the use of communications technology is especially appropriate. Therefore, such schools would be the best audience for examination and discussion of how technology can be made most useful.

When the questions about the use of new technologies are asked in the context of these types of instructional goals, then the technology has a much clearer role. As one panelist commented, "The use of technology, by itself, is not the goal; the goal is to examine the forms of assessment that we want to create and then determine how technology can help." He further stated that technology needs to be seen as fulfilling a need, "rather than be a solution searching for a question." Another panelist made a similar point, stating that first and foremost in looking at technology it is important to consider the impact on student learning, and determine assessment based on this.

4. Share efforts in the development of assessments using technology

Communications technology itself can help promote the development of new assessment tasks and standards by making possible communication about their efforts among persons in a range of geographic areas. Panelists mentioned the importance of sharing exemplars and standards among those who are developing assessment approaches, and of developing a consensus across sites on scales for judging performances or other work examples. Telecommunications technologies, by supporting this type of sharing, can thus help to hone common judgments and definitions of goals and standards. For example, the panelists referred to the value of communication via the Internet for teachers and others who are involved in the development of assessments.

One suggestion was to develop a collection of exemplars and ratings for those exemplars (e.g., utilizing World Wide Web) to build a consensus about assessment. Such a collection would assist others in development efforts while at the same time helping to build more consistency across schools and districts in the types of assessment standards that are developed.

Another panelist noted that the increased potential for sharing of assessment models and exemplars can make possible a more bottom-up as opposed to top-down process in developing standards. A school or local school district, using state guidelines, could develop

their own specific standards, and could provide examples of these to the district or state as a clear definition of the standards to which they should be held accountable.

D. What Factors Limit Use of Technology for Assessment?

The panelists mentioned several factors that can limit the effective use of communications technology and other technologies for assessment. These factors are quite varied and reveal that effectiveness is a function of an overall commitment to the assessment. The factors discussed include the following:

1. Lack of "fit" with the instructional setting

Although technology can become a valuable new tool for assessment, the panelists cautioned that it cannot change assessment by itself. There needs to be an environment in which the use of the technology is consistent with the goals and approaches used within instruction. As one panelist stated, with reference to the Digital Portfolio: "Put this tool in a traditional setting, and it simply becomes one more thing to do." Its use only makes sense within a school that is already utilizing alternative assessments.

Another panelist stated that a where instruction and assessment are focused on research, inquiry, and interpretation, rather than memorization, then structural supports within the program need to be provided. For example, the structures of time within the school need to allow for longer class periods and opportunities for teachers to meet, learn, and collaborate need to be built in. Without these, the instructional setting will limit the potential for effective use and integration of the technologies.

2. Lack of "fit" with the overall accountability system

One of the panelists observed that communications technologies will not be used as effective tools for assessment until the whole system of accountability is changed. She noted that there are schools and districts that are carrying out substantial reforms in instruction including integrated use of technologies, but that all of these are dropped out of sight when state-mandated tests come up. Then, when it is time for the standardized tests, all of the work on problem solving, extensive reading and research stop so that teachers and students can practice the rote skills needed to pass the test. Within this type of context, the panelist could not foresee communications technologies being used effectively for assessment.

3. Need for development of assessment models and methodologies

A major limitation as identified by one panelist is the fact that the kinds of assessment environments that have been discussed, e.g., placing students within a realistic situation to test language use, have primarily been developed and used for the purposes of teaching and less so for assessment. For this reason, the systems needed for scoring performances have

not been developed. There need to be new ways of evaluating performance in order to carry out this type of assessment effectively.

Thus, this panelist predicted that extensive work will be needed in development of appropriate new methodologies for the newer types of authentic testing, just as there were years of development involved in working with the psychometric-based tests, beginning from the time that Benet first developed his test. Development of new views of testing and appropriate methodologies will take time. Other panelists similarly pointed out the need for work on how to describe student skills and student learning within the new types of learning and assessment environments that communications technologies and other technologies make possible.

4. Barriers of access to communications technologies

One panelist described availability of access as a key barrier to use of communications technology for assessment. A recent study by the National Center for Education Statistics showed that only 3 percent of the nation's classrooms have access to the Internet. Teachers need time to experiment and explore the technology, and they need to have assistance at hand as they do so both for the hardware and software problems that will arise. At a minimum, there should be access for teachers from their classrooms, ideally, they should also have access from home. The panelist commented that research has shown that teachers learned considerably faster when they have access to technology from home.

Another panelist noted that as new technologies provide new opportunities, there will need to be some assurance that no learner is denied basic access to educational opportunities due to lack of a computer. However, he further commented that since there will always be disparity among learners in access to technology, it will be critical to ensure that a new strategy be designed to accommodate everything from the lowest to the highest technology. For example, he mentioned the fact that conversant technologies, either voice-driven or keypad-driven, can be used to provide feature-rich access to those without computer access.

A third panelist suggested that access to telecommunications must become an assumed tool for education, just as it has become for the work environment. This would mean that there must be a plan to have a computer available for every teacher and student so that they can make use of telecommunications whenever needed. Most state and school plans, even though recently more ambitious, do not reach for this goal.

5. Difficulty in implementing a standardized approach to technology

One panelist sees the educational institutions themselves as providing barriers to the development and implementation of new communications technology. For example, he commented that tensions between the K-12 and higher education communities will make it very difficult for educators to agree on a standardized approach to the use of technology. The development of agreement on standard technologies and on approaches in working with technology will be important to the development of its wide-spread availability and use.

6. Expense

The expenses involved in use of technology were noted as a major barrier to its widespread development and implementation. However, as was also noted, the cost of technology is likely to decrease as the systems become more common and easier to acquire. In addition, if the use of technology for assessment helps to reduce costs in some areas (e.g., use of distance-based assessors on an as needed basis) then more résources may be available to build further capacity in terms of technology. Other comments indicated that cost is still an issue, since there will be additional costs for a school or district to build in the capacity for use of communications technology, from basic equipment purchase to installation and management, and then the training and support for personnel.

E. Summary

As evident in the above discussion, the panelists' responses indicated uses of new technologies such as communications technology will offer new views and models of assessment. Three key points were made by the panelists:

1. Technology provides significant change in the nature of learning and assessment tasks

All of the panelists commented on the relationship between the nature of learning and assessment and the important role that technology can play in both. However, as one panelist commented, no technology is the solution to our questions, instead technologies should be viewed as "enablers and enhancers" which when used properly greatly expand the nature of tasks used for learning and assessment within the classroom.

2. Technology broadens resources for assessment

The panelists noted that the use of technology, especially communications technology, broadens the range of resources available for assessment. Access to resources is broadened in two ways. First, through use of technology, classrooms and schools will no longer be required to rely on what is available locally. Instead, they can utilize telecommunications (e.g., via satellite, modem, or network) to cross borders of time and place to gain access to the best tools and testers available. For example, outside assessors, various "stakeholders" such as state departments of education, higher education, or other groups could also participate in assessing student work.

Second, technology broadens resources by offering new aspects of students' work for observation. Examples mentioned were focuses on student processes in producing a work product, student interaction with others in the context of problem-solving activities, and opportunities to observe students' use of skills applied to realistic situations. Thus, technology broadens resources by offering these new ways of looking at students and at the skills they possess.

3. Technology offers opportunities for increased collaboration in development of assessments

The communications technologies opens up opportunities for persons who are developing assessments to join efforts and share progress. As mentioned earlier, assessment tasks, rating scales, standards, and exemplars of student work related to standards can be collected and shared via the Internet or other means. The end result will be the "'democratization' of access to high quality learning, assessment, and other opportunities." In this way, for example, there can be increased opportunity for those who are familiar with assessment of language minority students to provide input to assessment efforts across a range of areas.

4. The role of technology in assessment

Panelists commented on the role of communications technology and other technology within assessment. Two important perspectives on the role of technology were given. First, technology by itself cannot change assessment. Use of communications technology may make possible new ways of viewing students' learning, including development of higher order skills. However, if the classroom or school has not focused on these as indicators of student progress, the technology will not prove ultimately to be useful for assessment.

Second, the panelists emphasized that technology should be a tool to focus on student learning in new ways. Communications technology and other technologies offer tools to help educators to focus on student work in new ways, to help do things better, and perhaps differently. Technology should not be viewed as simply a faster and easier means of doing the same type of assessment we have been doing. Technology should be used to explore and reach beyond our existing approaches to assessment to include those that make possible new understandings of how students learn, how we can observe that learning, and how we can use what we learn through the new assessments to improve instruction for all students.

IV. Conclusions and Recommendations

The questions posed in the written focus group concerned the use of communications technologies to address assessment needs, and the implications of technology for assessment of limited English proficient students. In responding to the five questions, the panelists discussed a range of technologies, especially those that involve communications technology, interactive videodisc technology, computer-based portfolios, and simulations, with many of these combining use of video, audio, text and graphics.

The purpose of the written focus group was to consider communications technologies in particular, with a focus on future technology uses in assessment. Thus, the panelists' comments were focused on a range of newer technologies as opposed to other available technologies such as computer-based testing in more standard formats, e.g., involving multiple-choice responses, or approaches such as computer-adaptive testing. Thus, while many types of technology use could be included, the focus here was on exploring the implications of the newer models of technology use, and on discussing their possible implications for assessment of language minority and limited English proficient students.

The first section below presents the key points made by the panelists regarding the implications of communications technology and other technologies for assessment of limited English proficient students. The second section presents recommendations to researchers, evaluators, and practitioners and to OBEMLA regarding technology uses and assessment of limited English proficient students. The recommendations have been developed based on the findings reported in Chapter III.

A. Implications of Use of Technology for Assessment of Limited English Proficient Students

The panelists commented that use of communications technologies and other technologies offer new capabilities that can be of benefit for assessment of limited English proficient students in particular. They suggested that use of these technologies can offer the following:

1. Increased access to native language assessment resources

Use of distance-based assessors, whether via telephone or other media, can provide access to language proficiency assessment and academic assessment resources in the native language that would not otherwise be available.

2. Direct demonstration of student skills

Use of technology (e.g., computer-based portfolio assessment, simulated environments for problem-solving, or other) allows direct demonstration and observation of a student's work. For language minority students who enter school with differing levels of literacy skills, of

prior schooling, of skills in content areas, and in English language proficiency, this is an important benefit. Direct observation of samples of a student's work provide opportunities for teachers and other assessors to gain a clearer understanding of an individual student's capabilities and needs. Student work samples can be designed so that they are appropriate to the student's level of ability, perhaps offering exemplars of skills that might otherwise not be observed, while still being consistent with the goals established for the school.

3. Increased opportunities to track growth

The technologies that have been discussed offer the potential for more closely tracking student growth in skills, through comparison and observation of the processes students engage in as they work through to a solution or end product. This very individual and close look at the student's performance would allow for more finely tuned comparisons across time of student abilities. Focus on assessment of student growth in skills, as discussed in Hopstock* is one alternative to be considered as a part of an overall assessment plan for limited English proficient students.

4. Opportunities to demonstrate skills in a variety of media

Access to a variety of media (video, audio, graphics, text) broadens options for demonstrating specific skills. This is a particularly important benefit for those students whose lack of English language skills limits their participation in more standard forms of assessment.

B. Recommendations

The purpose of this section is to provide recommendations based on the panelists' comments. The recommendations are provided in two sets: recommendations for researchers/evaluators/practitioners and recommendations to OBEMLA.

1. Recommendations for Researchers/Evaluators/Practitioners:

(a) *Use technology for assessment only with a clear view of its purpose within the program*

Technology offers considerable potential for broadening the nature of skills that can be assessed and for providing a closer look at individual student performance. However, for an effective use of any assessment, the purpose

---

* Hopstock, P.J. (1995). *Recommendations on Student Outcome Variables for Limited English Proficient (LEP) Students.* Arlington, VA: Development Associates.

of the assessment, including the types of information to be obtained and the uses for that information within the program, should be clearly understood and should be consistent with the overall goals of the program.

(b) *Use communications technologies and other technologies for assessment as part of a collection of assessment measures*

The panelists focused on communications technologies and other technologies that present new capabilities for assessment. However, as noted in the findings, each different type of technology offers a different view of a student's abilities. Use of a range of different assessment measures will therefore provide the most comprehensive picture of an individual student's skills at any one time and of that student's growth over time.

(c) *Share with others using technology for assessment, and develop common standards, scales, and exemplars where possible*

Communications technologies can be used to support development of shared understandings about assessments. This exchange of experience and methodologies can help to build a shared system of assessment resources that will assist overall in the development of effective approaches to assessment.

(d) *Consider use of technology for assessing student processes in completing products or reaching solutions*

Assessment measures have most typically relied on evaluation of a final student product, whether a set of answers to test items, a finished text, or other product. The findings of the written focus group indicate use of technology to observe how a student reached the final answer or developed the final product, can greatly assist a teacher in planning instruction and may be a more sensitive indicator of student learning and growth.

2. **Recommendations for OBEMLA:**

(a) *Support training in the use of a variety of technologies, including communications technology, and their use in assessment as a component of training programs for teachers of limited English proficient students*

The findings indicated that teachers are a key component to the success of technology use and that it is important to include considerable staff development and assistance. A first key point at which to familiarize teachers with technology for instruction and for assessment is in their training programs. Thus, in its role of supporting programs that train teachers of limited English proficient students, OBEMLA should encourage programs to include training in the technologies available and in principles and practice for the effective integration of technology in their classroom.

(b) *Identify and promote the establishment of a network of programs that serve limited English proficient students and that use technologies for instruction and assessment*

There is a need to maximize what is being learned from those who are already utilizing technology for instruction and assessment with limited English proficient students. To promote development of knowledge about the use of technology, and in particular about its use with limited English proficient students, OBEMLA should carry out the following steps:

(1) Identify programs that work with limited English proficient students using technology;
(2) Identify the specific nature of the technology being used and how it is used;
(3) Establish mechanisms for programs working with similar technologies to share information (e.g., via the Internet, telephone, or other means); and
(4) Provide opportunities for summaries of the different approaches, methodologies, and exemplars. Telecommunications technologies could support this effort, e.g., using the World Wide Web as a location for such information.

(c) *Conduct research into the participation of limited English proficient students within use of communications technologies and other technologies*

As experience is gathered in the use of technology with limited English proficient students, OBEMLA should support more directed research into the use of specific methodologies and into the principles and practices related to their effectiveness for assessment of limited English proficient students. For example, investigations could focus on methodologies for assessing the performance of limited English proficient students within group problem-solving in use of a computer simulation, or for assessing their participation in scientific research tasks involving use of the Internet.

(d) *Identify resources and examine possible models for use of distance-based assessors of limited English proficient students to expand the resources available to schools and districts for language proficiency assessment and academic assessment*

The use of distance-based assessors could offer substantial new resources to schools and districts for assessment of limited English proficient students. In order to realize this potential, however, a number of steps would be needed. These would include the following:

(1) Identify the key types of language resources needed;
(2) Identify sources of expertise to address these needs (e.g., language assessment experts fluent in specific languages, content area experts with fluency in specific languages);

(3) Identify the specific types of technology through which the language resources would be best accessed (e.g., telephone only, telephone plus video, computer-based portfolio, other);
(4) Develop models/methodologies for conducting distance-based assessments using the specific technology(ies) identified;
(5) Develop mechanisms for creating the links between schools/districts with assessment needs and assessors;
(6) Develop means for obtaining agreement on approaches, standards, etc. among on-call assessors; and
(7) Provide mechanisms for maintaining the pool of on-call assessors.

In the findings, one recommendation offered was to consider development of alliances between business and education. There may be potential for exploring this type of alliance in developing on-call assessments.

(e) *Identify means by which equity in access to technology can be provided*

A key concern is that not all schools and not all students will have equal access to the additional resources technology provides. The findings reported here suggest that some alternative means of providing access via different technologies may be available (e.g, in some cases through use of voice- or keypad-driven "conversant" technologies). However, a range of solutions will need to be identified for offering increased access to computers for those who are without.

The issue of equity is of concern not only for access to computers and technology, but also for access to more challenging uses. Lower income and minority students are more often exposed to use of computer for tutorial uses* as opposed to more exploratory purposes which involve the use of higher order cognitive skills. This suggests again the need for development of further resources for use of technology with all students and the importance of providing training to teachers—in this case, to teachers of language minority and limited English proficient students in particular—in ways in which these students can be fully included in activities involving use of technology.

---

* Means, B., Blando, J., Olson, K., Middleton, T., Morocco, C.C., Remz, A.R., & Zorfass, J. (1993). *Using Technology to Support Education Reform.* Washington, DC: U.S. Government Printing Office.

Notes

1 The prior related reports are the following:

Zehler, A.M., Hopstock, P.J., Fleischman, H.L., & Greniuk, C. (1994). **Task Order D070 Report: An Examination of the Assessment of Limited English Proficient Youth**. Special Issues Analysis Center. Arlington, VA: Development Associates, Inc.

Zehler, A.M., Hopstock, P.J., DiCerbo, P.A., Heid, C., & von Glatz, A. (1995). **Literature Review and Synthesis Report on Institutional Change and Its Implications for Schools Serving LEP Students**. Special Issues Analysis Center. Arlington, VA: Development Associates, Inc.

Fleischman, H.L., DiCerbo, P.A., & Hopstock, P.J. (1995). **Research Designs for Measuring Institutional Change Affecting the Education of Limited English Proficient Students: Focus Group Report.** Special Issues Analysis Center. Arlington, VA: Development Associates, Inc.

Hopstock, P.J. (1995). **Recommendations on Student Outcome Variables for Limited English Proficient Students**. Special Issues Analysis Center. Arlington, VA: Development Associates, Inc.

2 Jeffrey Munks, personal communication, 1995.

Appendices

Appendix A: Focus Group Participants

Appendix B: List of Questions

Appendix C: Responses from the Participants

Appendix A:

List of Participants

SIAC Task Order D190 Focus Group Participants

Dr. Allan Collins
Northeastern University/
Bolt Beranek and Newman, Inc.
135 Cedar St.
Lexington, MA 02173

Dr. Margaret Honey
Center for Technology and Children
Educational Development Center
96 Morton Street
New York, NY 10014

Mr. Jeffrey J. Munks
10584 Hidden Mesa Place
Monterey, CA 93940

Mr. David Niguidula
Coalition for Essential Schools
One Davol Square
Providence, RI 02903

Appendix B:

List of Questions

SIAC Task Order D190 Questions for Panelists

1. What has been your experience (or other with which you are familiar) in the use of communications technologies in the (1) assessment of language proficiency and/or (2) assessment of academic skills? What specific technologies (e.g., distance learning, videodiscs, etc.) and student populations (e.g., deaf, limited English proficient) has this experience involved? Please include any experience with which you are familiar that involves co-location of the assessor and the student.

2. How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in (1) the assessment of language proficiency and/or (2) the assessment of academic skills?

3. In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?

4a. In your opinion, what specific communications technologies hold the greatest potential or promise for improving (1) assessment of language proficiency and/or (2) assessment of academic skills? Please explain your choice and discuss the ways in which assessment will be improved through the use of the technology(ies) you specify. For what student populations would these be most effective?

4b. What do you think are the limits of these particular communications technologies for the purpose of educational assessment?

4c. Does the use of technology lead to new models of assessment? What implications do these changes have for assessment of language minority and limited English proficient populations?

5a. What do you think are the future possibilities for developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

5b. What do you think are the potential difficulties involved in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

5c. What suggestions do you have for how to best take advantage of these future possibilities or to overcome potential difficulties?

Appendix C:

Responses from the Participants

1. **What has been your experience (or other with which you are familiar) in the use of communications technologies in the (1) assessment of language proficiency and/or (2) assessment of academic skills? What specific technologies (e.g., distance learning, videodiscs, etc.) and student populations (e.g., deaf, limited English proficient) has this experience involved? Please include any experience with which you are familiar that involves co-location of the assessor and the student.**

Response from: Allan Collins

I have no specifically direct experience on this issue. I've been involved in several projects to try to develop techniques using video and computers to assess students in their science learning. I have taught a course at Northwestern University on several occasions on performance and portfolio assessment.

One of my students here at Northwestern built a system that was designed for Spanish speakers to help them learn English. This was built by Enio Ohmaye. Basically the system interacted with people visually, so you would come into the airport at O'Hare in Chicago and you would talk to somebody who would direct you how to get to a place out west of Chicago. And then you arrive at the hotel and you would have to try to register and get a room and then there would be various events that would occur to you. There was support in the system to help you carry on conversations and carry out task. The inputs were all typed inputs. For the non-English-speaking student, the system had a lot of videos of people talking to you in English. You could also get help by seeing the words that they said in a typed form. You could get a translation of the typed form if you needed it. So there were various kinds of assistance that you could get in trying to interact in these rather naturalistic situations. So it was a system that was basically designed to teach English as a foreign language in the context of doing everyday tasks. The system could be used for assessment in the same way that it's used as a teaching system by scoring how well the student responds in each of the situations and how much help they need.

Another system that was built at Northwestern was designed to teach people how to speak to clients on the telephone. It was built for Ameritech. The system would record your voice and you could compare the speech-intonation-pattern line of your voice (e.g., intensity and inflection were recorded) to that of a model speaker. So basically the system was allowing you to see how well you responded in politeness terms in a situation where people were asking you for information and making complaints.

I should also mention that Bolt, Beranek, and Newman, where I work most of the year, did build a system for the deaf some years ago which would display the placement of your tongue for the vowel in a word that you spoke. So it would ask you to say a word like box or cat and it would show the position in the throat where the sound that you produced was as compared with where it should be for the actual spoken word. So again, that could be used to assess how well you positioned the vowel in the throat.

I also wrote a proposal once, which was not funded, to build an environment where non-native speakers would try to carry out tasks that were given to them. The instructions would be given in English and then they would have a task to do that required understanding what the instructions were. The task might be something like, "click on the box in the right hand corner". So you would measure their understanding of the spoken language by having them carry out various tasks that were simple or more difficult. And again, that kind of system could be used as an assessment device.

Finally, I had a student at Northwestern who tried to do visual analysis of a video movement. What he did was take videos of people vaulting. Since there are certain positions that the vaulter is supposed to maintain as they carry out a vault, the notion was to try to be able to do automatic scoring by analyzing whether the vaulter actually was in the positions during the vault that they should be. That is getting into the automatic scoring and you could obviously try to do automatic scoring of speech in similar ways. But automatic scoring is still very primitive.

I do have a paper on the uses of video and computers in assessment, which I will include in the package.

1. **What has been your experience (or other with which you are familiar) in the use of communications technologies in the (1) assessment of language proficiency and/or (2) assessment of academic skills? What specific technologies (e.g., distance learning, videodiscs, etc.) and student populations (e.g., deaf, limited English proficient) has this experience involved? Please include any experience with which you are familiar that involves co-location of the assessor and the student.**

Response from: Margaret Honey

For the past five years, the Center for Children and Technology has been involved in the Literacy Network Project. Housed at Lexington School for the Deaf in New York, this project uses a Local Area Network (LAN) to enhance subject matter learning and literacy development in deaf students. High School students have used this networked system of computers, equipped with communications software, in their science classes. Discussions and activities are conducted in written English over the network, and students have an opportunity to practice reading and writing as part of meaningful and purposive learning activities. The results of CCT's research indicate there is improvement in student's writing and thinking skills in those classrooms in which the network was used frequently and consistently (See the enclosed newsletter, Literacy and Technology).

In another project, CCT has been working with Bell Atlantic and the Union City Board of Education at a middle school in the District. The Union City Board of Education serves 8,541 students in eleven schools (three elementary, five K-8, one middle, two high schools). Approximately 91% of the students are Latino and 75% of these students do not speak English at home. Thirty-four percent of the students are enrolled in the District's bilingual program and over a third of the District's teachers are certified ESL or bilingual. The majority of residents are of low or moderate income, and 17% of the District's students have been in the country less than three years. In 1992, there were 2,537 Aid for Dependent Children households in Union City with 4,597 children. Seventy-nine percent of the District's students receive free or reduced price lunches a figure that is three times greater than the national average of 25.9%.

At the outset of the 1993-94 school year, Bell Atlantic supplied all Columbus School teachers, the school's principal and curriculum resource teacher, and all seventh grade students with 486-level computers with telecommunications capabilities at home and at school. In addition to the 160 workstations residing in teachers' and students' homes, 44 workstations were distributed throughout the school's classrooms. Lotus Notes is used as the communications platform and Microsoft Works and Publisher serve as basic software tools. Using the PCs at home and at school, students and teachers are able to access various remote servers, on-line CD-ROM resources and encyclopedias, and send e-mail locally and over the Internet.

Students scores on state-wide tests indicated that the prevalence of communications technologies in the school and in students homes is having a beneficial effect on student learning. The eighth grade students at the Columbus school were the only students in the district to meet state standards on New Jersey's Early Warning Test (EWT). In order to meet

state requirements, 75% of the students must pass in each of the three subject areas (reading, math, writing). Columbus students did better than this: 87.5% passed reading, 78.5% passed math, and 86.5% passed writing.

In a practice Early Warning Test administered to Union City seventh graders, Columbus students had the highest overall scores. They had the highest pass rate in math (58.6%) and in writing (69.3%), and finished third out of five schools in reading (65.8%). According to the Director of Academic Programs, the EWT Columbus writing scores, which range form 10% to 40% higher than other schools in the district, can partially be attributed to the amount of writing and editing that students are doing in Columbus's technology rich environment. (For more detail, see the enclosed report: Union City Interactive Multimedia Education Trial.)

1. **What has been your experience (or other with which you are familiar) in the use of communications technologies in the (1) assessment of language proficiency and/or (2) assessment of academic skills? What specific technologies (e.g., distance learning, videodiscs, etc.) and student populations (e.g., deaf, limited English proficient) has this experience involved? Please include any experience with which you are familiar that involves co-location of the assessor and the student.**

Response from: Jeffrey Munks

In building a nationwide network of telecommunications based interpreters (AT&T Language Line Services), we began using communications technologies for distance based testing and assessment in the middle 1980's. We contracted with an outside organization (ACTFL) to provide qualified raters who could conduct oral proficiency interviews (OPI's) over the telephone. We would schedule the exam in advance and then have the rater call the individual who had applied to work for us. The rater would, with the applicant's permission, record the exam and then apply a rating (based on the ACTFL scale). The tape recording and the rating would be sent to us. We would then send the tape to a second ACTFL rater and pay to have a blind rating done. If the two ratings matched and met our criteria, we would accept the applicant and then use the telephone and mail to begin the process of training the person to interpret for us. This approach has been used over the past ten years and has enabled us to build a force of interpreters which spans the continent and covers up to 140 languages 24 hours a day. Using a sophisticated combination of telephony and computing, the system provides an average interpreter connect time of 45 seconds. A variation of this model would be worth considering as this study moves forward (for educational applications).

All of the OPI's we conduct are administered over the telephone since that is the medium in which the successful applicant will be working. Administering the exam in a co-located, or face-to-face setting would change the dynamic of the exercise.

Additionally, ongoing training and professional development of interpreters is conducted over the telephone utilizing a digital conferencing bridge. Currently, as many as 39 students can join an instructor on a remotely managed conference call. Using written materials mailed out in advance of the session, students will work through a variety of instructor-led and peer-involved exercises. Typically, the sessions involve interpreters who carry the same combination of languages and work will be done in English and the target language.

The telephone based training is supplemented once a year with a three day conference held in Monterey. With the assistance of experts at the Monterey Institute of International Studies and other local resources, Language Line sponsors intensive training sessions which cover a broad range of subjects. The experience enables people who had known each other only by voice to connect a face and physical presence to their co-workers, classmates, teachers, supervisors, and others who they deal with over the phone on a regular basis.

1. **What has been your experience (or other with which you are familiar) in the use of communications technologies in the (1) assessment of language proficiency and/or (2) assessment of academic skills? What specific technologies (e.g., distance learning, videodiscs, etc.) and student populations (e.g., deaf, limited English proficient) has this experience involved? Please include any experience with which you are familiar that involves co-location of the assessor and the student.**

Response from: David Niguidula

My research over the past few years has examined a tool we call the Digital Portfolio. This is a multimedia tool used to record and organize student work. Currently, we are testing prototype software in six settings: four high schools, one middle school, and one elementary school. Of these, one high school is in a major city, serving a primarily minority population.

The Digital Portfolio is software that runs on IBM and compatible machines with Windows. In the portfolio, students can store their work, once they have put it in a digital form. That is, the portfolio can handle text, graphics, audio, and video, but that information must be typed, scanned, or digitized by the student.

The software reflects an overall strategy for school reform. A school has to ask itself, "what do we want our students to be able to know and do?" Answering this question presents a *vision* of what qualities a graduate should possess. From there, the faculty needs to tackle how students can *exhibit* those qualities -- what specific tasks a student can perform in order to demonstrate that he or she has the skills and knowledge that fulfill the vision. The third question then becomes, "How do we arrange our systems so that all students can complete these exhibitions?" That is, from this vision, how should the curriculum, scheduling, and so on, be arranged so that exhibitions can be successfully accomplished by all students?

One such system is the use of technology, which needs to be deployed in the service of helping students achieve the school's vision. The Digital Portfolio may be helpful to schools in keeping the vision a living statement, as opposed to a document created for review purposes at accreditation time and filed away for the next ten years. The "main menu" of the Digital Portfolio (see Figure 1) represent the vision of the school. (This menu could be different for every school.) Thus, when a teacher (or any reader of the portfolio) wants to review student work, it is organized by the components of the vision; similarly, when a student decides that something is a "good" piece of work, he or she has to determine (with a teacher, typically) what parts of the vision are represented by the work.

For each entry, students enter their work, and the goals, or components of the vision, that the work represents; in addition, each portfolio entry includes the assignment distributed by the teacher, and evaluations by teachers, students, and/or outside judges (see Figures 2 and 3). This point of multiple evaluations is critical; it says that there is not one correct way to look at a piece of work, but multiple ways, depending on what it is that an assessor wants to find.

The entire point of the Digital Portfolio is to allow a student to present a richer picture of his or her abilities than traditional assessment records, like report cards or quiz scores, can show. The key thing is that any evaluator of the students' abilities can look at actual work, rather than the abstractions of letters and numbers.

Now, our prototype is not the only such tool that exists; Scholastic's *Electronic Portfolio* and Aurbach and Associate's *Grady Profile* are also tools for collecting student work in multimedia software. What is missing in those tools, we believe, is the definition of a vision, and thus our software presents a different organization on the same set of data: an organization that should help readers of the portfolio more quickly determine if a student has the abilities that the reader wants to see.

There are (at least) three assessments involved with the Digital Portfolio, or just about any form of portfolio assessment. First, typically, a student's work is assessed in the context of a course; that is, a project is completed in Algebra or U.S. History, and it is evaluated by the teacher. Second, a student (and, often, teacher) determines if this piece of work is a good representation of his or her abilities of some component of the vision. Thus, a student might wait until the end of the year, and examine all of his or her work and determine which should be in the portfolio; more commonly, though, students enter items into their portfolio as they go along, judging if a particular piece has enough merit to become part of the portfolio. Finally, an outside reader, be it a parent, or a college admissions officer, or the following year's teacher, assesses the student work.

Language proficiency can certainly be one component of a school's vision of what a graduate should be able to know and do, and the Digital Portfolio can allow a student to demonstrate that ability using whatever medium is appropriate, depending on whether the demonstration requires the printed word or the spoken word; casual conversation in class, or a formal presentation of an idea. The point, again, is that the vision drives the assessment; the expectation is that the student will meet the goal in whatever form makes sense, and that the Digital Portfolio allows that work to be stored and examined by future readers.

I'm not sure I know what "co-location" means. I take it to mean that the assessor and the student are in the same space. Our work does not depend on all assessors and students meeting face-to-face, but in the schools' experiences with implementing digital portfolios, every student needs some adult in the school to be concerned with his or her portfolio. That is, the student needs to be able to talk to someone about what should go in the portfolio, and what represents "good" work.

2. **How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in (1) the assessment of language proficiency and/or (2) the assessment of academic skills?**

Response from: Allan Collins

Well, I guess I see three kinds of roles here. One is in creating environments or contexts in which language or other kinds of tasks are carried out. The second role is actually recording the performances: you could record in video, or in audio, or on a computer what the student did. That's a recording of the performance. And the third aspect that technology might be used for is in actually analyzing or scoring the performances that occur.

So let me talk about each of these roles for technologies. The context setting or the posing of problems or questions to the person being assessed could either be done by a human at the end of a line, so that you could have assessor in one place and the person being assessed in another place. Alternatively, you can have the problems posed, or situations you put the person in embodied in a computer kind of environment. So when I talked about the program where you come into an airport and then go to a hotel to register, those are creating situations in a computer system and that material could be sent out to different places as an assessment device. So that allows you to create fairly realistic situations, to put persons into situations where it's interactive rather than paper and pencil which is not interactive.

The recording issue is discussed in the paper I'm enclosing. Video allows you to record different things than paper and pencil allow you to record, or typing allows you to record. It allows you to record gestures, it allows you to record ability to do hands-on experiments in science; it allows you to record how well one listens to somebody and asks questions. So video gives you one kind of window on performance—a different kind of window than paper and pencil. A third kind of window is a computer environment where you have to carry out realistic tasks, as I described, so we can see how well you do on different kinds of tasks. And the tasks can be very complex or they can be very simple.

On the scoring issue, we can only really score video, or spoken English by having judges rate it, much as judges rate written performances in a holistic or primary trait scoring scheme. With a computer task environment, where you're carrying out tasks, it would be perfectly easy to score how well you carry out those tasks. It also would be perfectly easy to score how much you improve if you're in a task environment for a long time, which is a measure of learning as opposed to just performance.

Another thing that is possible to do when scoring in a computer environment is to score how much help or hints you need in order to accomplish a task. This is related to a notion of dynamic testing that Joe Campione and Ann Brown have written extensively about.

2. **How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in (1) the assessment of language proficiency and/or (2) the assessment of academic skills?**

Response from: Margaret Honey

The strongest argument for using communications technologies for assessment purposes is that the technology -- when well designed -- can support students who are at very different academic levels. In other words, technologies can support flexible use. The Word Wide Web, for example, provides users with information in a variety of media (text, audio, graphics). The user can browse, read and view materials following their own preferences and interests, rather than being limited by the linear organization of traditional texts. How students choose to explore and conduct research in an environment like the World Wide Web, can be guided by their own level of expertise and understanding. However, if educators are to use the Web effectively with their students much more research is needed about the kinds of search strategies students of varying ability levels use on the Web. There is some research that suggests that when students are conducting research and taking notes using multimedia materials, they engage in more integrated and interpretive note-taking.

Multimedia composing or authoring tools also hold great promise as environments in which to authentically assess students work. Multimedia authoring tools supply a rich context for writing activities. Images, graphics and video can serve as prompts for generating text, and help students to express what may be difficult to put into words. The idea of producing for an audience is also very compelling to students, adding authenticity and value to their work. Again, more research is needed on student authoring in an environment like the Web. Teachers report that students are highly motivated and enthusiastic when undertaking Web authoring projects, but very little is known about the kind of learning that takes place or the parameters that need to be established so that teachers can make judgments about their students' learning.

2. **How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in (1) the assessment of language proficiency and/or (2) the assessment of academic skills?**

Response from: Jeffrey Munks

Much of what has already been written and talked about in terms of 'next generation' technologies is all around us but is simply not being used. Telemedia (or video telephony) is here and requires only ISDN (available from most telephone companies for a modest price) in order to activate. Paradyne Corporation makes a $300 modem that splits standard twisted pair copper telephone lines and makes it possible for me to call you on the phone, have you turn on your computer, and share word processing and/or data files while we are talking on the phone. This 'off-the-shelf' technology has tremendous potential for distance based language proficiency assessment and/or the assessment of academic skills. Imagine the following scenario: I call you on the phone and ask you to turn on your computer. You do but you are concerned because I am working in Word Perfect and you are working in Microsoft Word. Not to worry. I can export my program to your computer so you are looking at and working with a Word Perfect file. I tell you that I am going to write a sentence in Spanish on the screen and then I want you to read it out loud to me and then tell me what it means in English. Next, I am going to write a Spanish verb on the screen and I want you to write the conjugated forms of the verb right below it. The foregoing scenario is possible now with only a new external modem attached to existing hardware. The limitation here is that the communication is one-on-one. Other options are available, however. Private online service providers such as America Online (R) have already provided the capability to create private 'classrooms' on the network where 20 or more people can converse via computer in a free form or moderated session without regard to the traditional barriers of time or distance. Such forums can be a wonderful supplement to the kinds of exercises described above as well as the more traditional educational approaches. Of course, these and other distance based scenarios assume that all involved learners will have access to the technology required to participate. Obviously, there will always be a measure of disparity between and among learners (and teachers, for that matter) based on socio-economic issues, geographic location, etc. For these reasons, it is critical that any strategy employed be designed to accommodate everything from the lowest to the highest technology. As new technologies afford new opportunities, no learner should be denied basic access to educational opportunities simply for lack of a computer. Technologies such as conversant (which can be voice or telephone keypad driven) can do much to enable feature rich access for the technologically disadvantaged. Further, the notions of ubiquitous hardware and software, ease of upgradability, and technointeroperability should be given careful consideration at every turn.

2. **How can communications technologies, such as features of distance learning, the electronic transmission of video, audio, text, and graphics, be used in (1) the assessment of language proficiency and/or (2) the assessment of academic skills?**

Response from: David Niguidula

Allan Collins (Bolt, Beranek and Newman), Jan Hawkins (Center for Children and Technology), and John Frederiksen (Educational Testing Service) wrote a paper several years ago outlining how different technologies: paper and pencil, video, and computers, each provide a different picture of a student's capabilities. None is complete, but collectively, they provide useful information for assessing a student's abilities.

The technology cannot do the assessing by itself -- the sophistication of software is not far enough along to interpret language and analyze it in the ways we would want. At this point, the role of the technology is to record and organize assessments of student work. If a student can transmit his work to an assessor via satellite, modem, or network, then we have used the technology to create a link between student and assessor that wasn't previously possible.

The technology opens up the potential pool of assessors to anyone who is available electronically, which means that outside assessors from various "stakeholders," such as state departments, higher education, or other groups, could participate in assessing student work.

3. **In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?**

Response from: Allan Collins

The major way, is in the ability to look at their actual spoken language, and also to be able to see how they deal with many different kinds of situations, because you can put them into novel situations, and ask them questions or follow-up questions. So, what are the major ways that would improve effectiveness. It's in extending the range of your assessment so that you're not just looking at how people write answers to questions or write an essay or something like that. You can look at the spoken language and you can put them into realistic situations and you can look how they deal interactively with situations. And so those are the major effectiveness gains for less commonly spoken languages.

3. **In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?**

Response from: Margaret Honey

The power of communications technologies lies in their ability to provide a record of students work over time. In order for this to serve as a vehicle for assessment, however, teachers must be trained and supported in making judgments about student work. While this requires extensive professional development, the benefits to learning can be enormous. The state-wide experiments underway in Kentucky and Vermont are prominent examples of this -- although technology has not been extensively integrated into any of these experiments. Ideally, technology-based tools will be built that can support teachers and students in assessment and evaluation practices. David Niguidula's work on digital portfolios at the Annenberg Institute for School Reform represents an important step in this direction. How you determine cost-effectiveness in relation to any of this is complex. I believe that first and foremost one must consider the impact on student learning.

3. **In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?**

Response from: Jeffrey Munks

The imagination is the only barrier in considering answers to this question. For example, with the work the Soros Foundation is currently doing in the states of the former Soviet Union, it should soon be possible to engage a scholar in the Ukraine, via pc based telemedia, to conduct an assessment of the language (or any other skills) of a 14 year old recent arrival immigrant who needs to be placed in an appropriate grade level.

The cost implications of these technologies are potentially enormous. Imagine the notion of 'just-in-time' assessment. Rather than having locate someone qualified to do assessments and then match schedules and make arrangements for travel, lodging, etc., you could simply engage video dial tone, call the individual, and conduct the assessment almost ad-hoc. Such a scenario begs consideration of a new paradigm to describe cost and value based pricing. You might pay only for time-on-line with the assessment specialist. Such a specialist could be working from home and could be providing service on a global basis via video telephony. With this approach, the number of speakers of a given less commonly encountered language in a specific location becomes less of an issue because their needs can be addressed almost without regard to the 'clustering' traditionally required to achieve economies of scale.

3. **In what ways do you anticipate that communications technologies, such as features of distance learning, would improve the effectiveness and cost effectiveness of educational assessment of speakers of less commonly spoken languages?**

Response from: David Niguidula

The "effectiveness" of assessment depends on the type of assessment one wants to make. The Office of Technology Assessment's 1992 report *Asking the Right Questions: Testing in American Schools* addressed this issue, and the problems that result from taking one assessment and assuming it is a measure of something else.

In the school, the primary point of most assessments should be to evaluate a student's current understanding, thus indicating what next steps are appropriate in his or her education. In short, the teacher has to get to know the student well. Our work with Digital Portfolios assumes that a record of student work can be of some help to a teacher trying to get to know a student, but it may be that the mere process of *collecting* the work for a portfolio can help a teacher understand a student's abilities.

The Digital Portfolio may be particularly useful at the transitions: from one grade to the next, from elementary to middle, or middle to high school; from one setting to another when a student transfers. A record of the student's actual work, as opposed to test scores and letter grades, can help the new teacher or school understand where this student should fit into the school's system and curriculum.

Similarly, features of distance learning can bring groups together that may not be able to physically get together, allowing multiple assessors to evaluate a student's abilities. These multiple perspectives may be helpful in understanding the student's current capabilities.

Assessment of groups of students are more problematic. Again, one must ask why such assessments are done. There is a common assumption that schools have to be ranked against one another. While there are undoubtedly qualitative differences among schools, and certainly some schools that are failing our students, to rank all schools along a few measures doesn't make much sense. To me, then, an "effective" assessment of schools points out the schools' strengths and weaknesses for particular students or (perhaps) groups of students.

The technology can help by showing the process as well as the product. We do not have to evaluate an educational innovation for increasing language proficiency by looking just at the final test scores. We can record the activities in the classroom via videotape and multimedia and look at how goals are being met. We can see where the students began, and what steps they took along the way. Exactly which pieces of information will be useful will, again, depend on the goal of the assessment; communication technologies, however, can allow us to have a broader set of information from which we can assess a program.

4a. In your opinion, what specific communications technologies hold the greatest potential or promise for improving (1) assessment of language proficiency and/or (2) assessment of academic skills? Please explain your choice and discuss the ways in which assessment will be improved through the use of the technology(ies) you specify. For what student populations would these be most effective?

Response from: Allan Collins

With respect to the assessment of language proficiency, I would say that interactive video is the technology that has the biggest payoff, because you can put people into situations where they have to respond appropriately. They can respond either verbally to you through typed responses, or through carrying out some actions that you specify. The reason why this is a good idea is because it allows you to look at language proficiency in action, as it were. It's just a very different kind of language proficiency than a paper and pencil test can assess.

By making a single system that embodies this, you get an economy of scale that you can use this assessment device for lots of different situations, and for different populations. The populations it's most effective for are people who are trying to speak a foreign language. The other thing to say about the efficiency is that the system can put them into a variety of situations, and so there's no way you can quite prepare for it. You don't have to have items as they do in most testing now, but you record automatically how they deal with different contexts.

For assessment of academic skills, I think the biggest win is to put people into contexts where they have difficult problems to solve: troubleshooting an electronic circuit, dealing with a difficult client, or manipulating a physical object like a dynaturtle where you test their understanding of Newton's laws in different contexts. Maybe I should explain that. There is a system called Thinker Tools that allows you to give students activities to carry out and you need to understand Newton's laws in order to be able to carry out those activities. The point is that you can put people into situations where they have to solve difficult problems that require academic skills, and so you can then measure how well they do that.

4a. **In your opinion, what specific communications technologies hold the greatest potential or promise for improving (1) assessment of language proficiency and/or (2) assessment of academic skills? Please explain your choice and discuss the ways in which assessment will be improved through the use of the technology(ies) you specify. For what student populations would these be most effective?**

Response from: Margaret Honey

CCT's research has shown the power and potential of text-based communications technology with hearing impaired populations and with students of limited English proficiency. Assessment cannot be improved unless learning is improved, and our research shows that extensive opportunities to communicate with others on-line is not only motivating, but improves students' writing and reading abilities (see #1 above).

4a. In your opinion, what specific communications technologies hold the greatest potential or promise for improving (1) assessment of language proficiency and/or (2) assessment of academic skills? Please explain your choice and discuss the ways in which assessment will be improved through the use of the technology(ies) you specify. For what student populations would these be most effective?

Response from: Jeffrey Munks

I think that the ideal system to address the question is also the system that should be phased into K through University as language labs are upgraded across the United States. It starts with a basic platform such as Pentium based 486 PC workstation (very easy to upgrade as higher performance chips become available) equipped with a flex-camera, sound and video boards, head-set and whisper microphone. X numbers of these stations can be installed in a learning center. Individually, they can be placed in homes for not too much more than the cost of a high performance pc. In the lab, they are networked to a multi-tasking LAN that will support video, CD-ROM, and massive file storage capacity (via a large server). Mated to a switch, the system will allow ingress/egress so that students on-site can get onto the World-Wide-Web (WWW) and so that learners at home can dial in and gain access to the programs resident on the LAN. In keeping with the commitment expressed in the previous response, it would be appropriate to work a conversant technology into the mix so that learners without access to computers could use the keypad of their telephone and their voice to interact with the system. Total cost of such a system installed in a school based language lab (assuming 10 learner workstations) would be around $120k. Learners could access such a system from home with a 486pc with about $1k of upgrades. A Mac solution would be just as easy to craft. I would think the process of assessment could be dramatically improved through the addition of these types of technologies to the mix. The process would no longer be dependent upon people's ability to physically come together. Assessment could be done on an 'as needed' or 'when ready' basis and could be conducted in environments that are selected for specific effect.

4a. In your opinion, what specific communications technologies hold the greatest potential or promise for improving (1) assessment of language proficiency and/or (2) assessment of academic skills? Please explain your choice and discuss the ways in which assessment will be improved through the use of the technology(ies) you specify. For what student populations would these be most effective?

Response from: David Niguidula

Clearly, we believe in the potential of multimedia software as a tool for assessment. Allowing students to record work in whatever medium, yet having it in a convenient location, will allow individual assessors to look at the actual student work, and make judgments based on the work, rather than solely on the judgments of others.

An extension of our current Digital Portfolio would take advantage of the World Wide Web. The use of networked technologies adds the possibilities of assessors and students sharing ideas anywhere on the Internet. Now, we believe that the most important assessors are those that are closest to the student: the parents and teachers, and thus, putting a digital portfolio on a local area network is more important than placing it in some Internet-accessible form. Still, outside observers ranging from state departments to college admissions/placement officers would be able to get information about a student in an easily accessible and manageable form.

We also assume that students will select the work they feel best represents their abilities. For those students who are not still proficient in English, the use of a portfolio system allows them to show other skills that they have mastered.

4b. What do you think are the limits of these particular communications technologies for the purpose of educational assessment?

Response from: Allan Collins

The major limitation is that we have not developed these kinds of assessment environments yet. We've developed most of the kinds of systems I'm talking about for the purposes of teaching, so we haven't developed means for scoring performances very systematically. You can certainly score them automatically if it's a computer task environment. But there has to be a whole new kind of way of evaluating performance developed in order to carry this out. There also is the concern about the costs of equipment to do this.

4b. What do you think are the limits of these particular communications technologies for the purpose of educational assessment?

Response from: Margaret Honey

The limits surrounding the use of any technology for learning or assessment reside in the teachers ability to use them effectively. On-going staff development and training are essential.

4b. What do you think are the limits of these particular communications technologies for the purpose of educational assessment?

Response from: Jeffrey Munks

I think the biggest risk (or downside) to the use of existing and emerging technologies lies in their uninformed use and the attendant unrealistic expectation that so often accompanies such use. For too many teachers in an educational setting, new technologies are like hand grenades tossed into the classroom - you know something is going to happen when it goes off and it's never good... None of the technologies discussed here or in your other submissions represent solutions. They are, rather, enablers and enhancers which, when used properly, allow any or all of the following:

- Teachers can bring more and more creative learning opportunities into the classroom.

- Teachers can avail themselves of more data, more knowledge sources, more interactive exercise options for use in the classroom.

- Teachers can provide coursework and learning opportunities for distance based learners who either cannot make it into class or do not have access to traditional learning settings.

- Learners can engage the learning process anytime, anywhere, through a variety of access methodologies that range from standard twisted pair copper phone lines to and through satellite based schemes.

- Learners can expand their knowledge/information access options on a geometric scale and, in so doing, free themselves from over dependence on one source (the teacher).

- Language Learners, in particular, have the opportunity for a virtual total immersion experience by using the technologies previously discussed to bring native speakers into the home (through video telephony), realia into the home, dynamic interactive lessons into the home, etc. on an ad hoc or scheduled basis.

Again, the biggest threat in all this lies in the original introduction and orientation. All too often, teachers and students alike are provided with the technology and then told to go use it. The historical response to that approach has often found the teachers parking the technology in the corner in favor of the chalk board. The students, absent informed techno-guidance, invariably find a way to play games with the technology. Some schools have cracked the code by training their teachers up on techno-skills before introducing new technologies into the learning setting. I think it is a good step in the right direction. I also think there are probably a number of creative ways that particular process could be

accelerated. For example, what would happen if an entire school district committed to the purchase of a particular company's computer hardware with the condition that the company must first demonstrate its commitment to a long term relationship by providing training (at no cost) to every teacher in the district? If I were that company, I would be inclined to sit down with district administrators to work out a schedule so my best people could conduct seminars at each school site as soon as possible. I would also establish a toll free help line that would be available to teachers (and students) throughout the school year so that questions about the technology could be answered at any time.

4b. What do you think are the limits of these particular communications technologies for the purpose of educational assessment?

Response from: David Niguidula

The technology, by itself, will not change assessment. The Digital Portfolio, we believe, will not be useful in an environment where the school has not embraced alternative assessments already. Put this tool in a traditional setting, and it simply becomes one more thing to do.

Also, telecommunications of any kind cannot replace human contact. It is an open question as to how much people can get to know each other over wires or airwaves. Still, if a student has a relationship with another individual who may be outside the school building, telecommunications decreases the distance between the student and that assessor, and allows for more informed discussions about the student's abilities.

4c. Does the use of technology lead to new models of assessment? What implications do these changes have for assessment of language minority and limited English proficient populations?

Response from: Allan Collins

Yes, I think that technology is going to change the whole way we think about assessment. I think we will stop thinking about assessment as generating a set of individual items and rather think of putting people into realistic task contexts where they have to solve difficult problems or answer questions or carry out commands.

In assessing language minority and limited English proficient populations, we can put them into situations where we emphasize the critical things they need to learn to function in society. Then we can assess directly how well they function on the fly, as it were, interactively in a verbal context. And that's not something we could do with other technologies.

4c. Does the use of technology lead to new models of assessment? What implications do these changes have for assessment of language minority and limited English proficient populations?

Response from: Margaret Honey

The advantage of communications technologies, particularly the Internet and the World Wide Web, are that they support much more authentic learning practices. Students, for example, can use the Internet to engage in real-world science experiments, they can participate in cultural exchanges, or communicate in different languages. All of these experiences necessitate that different kinds of assessments be used to document learning. Multiple choice tests (the CATS or other standard measures of student achievement) do not do justice to the complexity of thinking and learning that take place in communications-based environments. (Please see the enclosed newsletter on Alternative Assessment and Technology)

4c. Does the use of technology lead to new models of assessment? What implications do these changes have for assessment of language minority and limited English proficient populations?

Response from: Jeffrey Munks

The use of technology will certainly lead to the development of new models for assessment. It is already doing so (the Language Line model previously cited). These changes mean that language minority and limited English proficient populations will no longer be dependent on whatever is available locally in the way of assessment tools and providers. Instead, they will be able to utilize the best tools and testers available. At the same time, the referenced technologies will enable those who produce testing and assessment instruments to work collaboratively across the barriers previously imposed by time and distance in an effort to ensure a standardized, quality based approach to the process. The end result will be the 'democratization' of access to high quality learning, assessment, and other opportunities.

4c. Does the use of technology lead to new models of assessment? What implications do these changes have for assessment of language minority and limited English proficient populations?

Response from: David Niguidula

The new model of assessment that this technology provides is one that completely focuses on student work, rather than its abstractions. While paper portfolios exist, the technology allows for an organization and communication of information that makes the information much easier to work with.

By focusing on student work, one is less tempted to rely on meaningless statistics, such as Grade Point Average, since it is harder to aggregate or average multiple pieces of work. This is not to say that we eradicate summaries; still, we can move toward qualitative, rather than quantitative descriptors of a student's abilities.

The use of technology also makes it easier to think of work as work in progress. As we see versions of work being created, we can see assessment as part of a feedback loop: a student creates, a teacher assesses, a student revises, a teacher assesses again, and so on. We can now track this entire progress. While we wouldn't want to see every draft of every piece of work, having such information available can help students and teachers examine the processes that the students use to create work.

The implications for language minority students, and indeed for all students, is that we can use technology to better understand each student as an individual, and examine his or her abilities on their own merits.

5a. What do you think are the future possibilities for developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: Allan Collins

Again, the potential is that you develop computer-based systems, some using interactive video, some not, which put students in contexts where they have to use spoken language, where they have to carry out tasks, following commands, and then assess them in those contexts.

5a. What do you think are the future possibilities for developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

5b. What do you think are the potential difficulties involved in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: Margaret Honey

These are complex questions.

First, there are barriers of access to communications technologies. According to a recent study by the National Center for Education statistics, only 3% of the nation's classrooms have access to the Internet (access meaning Internet email). Rural and urban schools are likely to have the least access. Teachers also need access to the technologies so that they have opportunities for experimentation and exploration. Our research has shown that when teachers have access from home, the learning curve is considerably shortened. At minimum, teachers need access from their own classrooms.

Second, there is the problem of time and training. It takes time and opportunities for professional development must be plentiful if teachers are to learn how to effectively integrate technology into their curriculum. Ideally, there is a school-based person who is free during the day to work directly with teachers in their own classrooms.

Third, there is the problem of accountability. Communications technologies will not be used as effective tools for assessment until our system of accountability changes. We have seen schools and districts that have brought about substantial educational reforms and done an excellent job of integrating technologies into these, "drop everything" when it comes time to take the state mandated tests. In other words, the research work, the extensive reading, and the sustained problem solving that students typically engaged in all stop and teachers work on practice exams and rote skills that are required to pass the test. Until this system changes, we will not see communications technologies used effectively for assessment purposes.

5a. What do you think are the future possibilities for developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: Jeffrey Munks

I think the future possibilities for developing and implementing communications technologies in educational institutions are excellent. They are also inevitable. To do anything other than weave the latest technologies deeply into the fabric of our educational systems is to risk seeing those institutions lose their societal relevance. In the area of assessment, tremendous advantage will redound to the benefit of those institutions with technological competence. They will be able to avail themselves of the very best tools and professionals and, in the process, substantially reduce 'time on task' in ways that will make them more cost and time efficient.

5a. What do you think are the future possibilities for developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: David Niguidula

This question goes beyond the issue of "assessment" and to that of "accountability." (The terms are used differently within the education world; for purposes of this document, I've been assuming that "assessment" refers to the evaluation of a particular student or group of students, while "accountability" refers to evaluation of schools as a whole.)

The key question in accountability is that of standards. The political debate has focused on the need for national standards. Yes, we want all American students to achieve high standards -- but who should set what those standards are?

It is our contention that technology can allow us to think about standards differently. Rather than assuming a top-down approach, where the federal government or some other body says, "these are the standards," we see national guidelines. Schools can become the focus of standards-setting; as communities, faculty, students, parents, administrators, and other interested parties can collectively answer the question, "What do we want our students to know and be able to do?"

Telecommunications provides an opportunity for a school to then take that question to the next step. In effect, a school can say to a state, "Here are our standards. You can approve them or suggest modifications, but once we agree on the standards for our school, you can hold us accountable to whether our students meet OUR standards." Because telecommunications can allow state or district personnel to visit a school, or to sample student work, without a physical visit, schools can make their work visible (and thus hold itself accountable) in an entirely new way.

Now, schools do need to "tune" their standards (the term "tuning," developed by Joe McDonald at the Annenberg Institute for School Reform, is similar to the tuning of a musical instrument: one compares a sample from one's own instrument with a sample from some outside source, and adjusts one's instrument accordingly). We think that schools can help each other to tune standards. Thus, what we see as a possibility are for schools to work together as "critical friends." If Schools A, B, and C are in a cluster, then a team from A can present its work to folks from B and C. The response from B and C might be, "We understand that you want all students to achieve a particular level of mastery of writing across the curriculum. But is the exhibition you have presented truly getting at that skill?" The schools get to know each other well enough to have that kind of conversation. The goal is to help each school tune its own standards -- not necessarily to have schools come up with identical standards.

Technology makes a great deal of this possible. Being able to send multimedia packages over the net allows schools to show what they are doing, and what student work and individual classroom settings look like.

Taken to an extreme, telecommunications could create entirely new structures. If a high school diploma is truly based on the exhibition of what we want students to be able to know and do, do we need those exhibitions to take place inside a school? The idea of a virtual learning environment, or more precisely, a number of environments where learning take place (ranging from schools to museums to businesses to workplaces) can be where a student spends his or her time learning; when he or she is ready to exhibit mastery, the student can make his work available to a school or other panel of judges via telecommunications. Now, I hope that not too much is read into this; I am not advocating the elimination of school, nor the idea that all students should be running all over the community with no purpose. We need structures for kids where there are adults who care about their intellectual growth. But, the possibilities of technology and new assessment systems can be a good reason for us to question our assumptions about what school is and could be.

5b. What do you think are the potential difficulties involved in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: Allan Collins

The difficulties are developing a new view of what testing is. The psychometric-based testing has taken a number of years to develop from when Benet first developed his test. There were years and years of methodology development. We need a similar kind of methodology development if we're going to have these more authentic kinds of tests. Then finally, we would need the technology to administer those kinds of assessments.

5b. What do you think are the potential difficulties involved in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: Jeffrey Munks

The major barriers to developing and implementing new communications technologies will come from within the educational institutions themselves. The academic community, dealing only with itself, often has a hard time deciding on a course and then striking out in pursuit of whatever course has been determined. The business community, on the other hand, is moving quickly toward the 'just-in-time' style of operations. This approach does not mesh well with the agonizingly slow process of 'deferred decision by committee' which characterizes so many educational institutions. Additionally, despite the rhetoric describing the importance of collaboration and cooperation between and among the various levels of education (and particularly so with the language folks) there is still a tremendous amount of tension between the k-12 and the higher ed. groups. This tension makes it that much more difficult to agree on anything remotely resembling a standardized approach to technology and related issues. Expense is the other major barrier to widespread development and implementation. On the good side, we see signs that the cost of technology is going down on a fairly rapid basis. As the systems become more ubiquitous and easier to get a hold of, cost will come down that much faster.

5b. What do you think are the potential difficulties involved in developing and implementing communications technologies for the purpose of educational assessment in schools, school districts, and state education agencies?

Response from: David Niguidula

Margaret Honey and Andres Henriquez's report *Telecommunications and K-12 Educators: Findings from a National Survey* (1993: Center for Technology in Education, Bank Street College of Education; for further information, contact Dr. Honey at the Center for Children and Technology, Education Development Center, 96 Morton Street, 7th Floor, New York) outlines many of the key issues in implementing telecommunications technologies in schools. Access to resources -- time, money, professional development, and, of course, equipment and phone lines -- are critical barriers to the use of telecommunications in schools. Most school and state plans for telecommunications, though much more ambitious than previously imagined, still do not assume a computer for every teacher or for any student whenever he or she wants to use one. The initial issue is to make telecommunications an assumed tool for educational endeavors (as it currently is to be an assumed tool for many work-related endeavors).

But let's assume that the infrastructure is in place to allow students and teachers to communicate electronically with anyone they choose. So what?

What are needed are compelling reasons for students and teachers to use telecommunications technologies. A compelling reason is not, "this will be useful later in life." A compelling reason might be, "this is something we can use for our current educational work."

Thus, developers need to understand what schools *can* be, rather than what they are now. We do not need tools to automate our current assessment practices; rather, we need tools that allow us to focus on student work in new ways. Telecommunications should be a tool for helping us do things better, and perhaps differently -- not just faster.

5c. What suggestions do you have for how to best take advantage of these future possibilities or to overcome potential difficulties?

Response from: Allan Collins

I think that we need to begin to develop methodologies for assessment that use technologies. We are just at the beginning of that kind of venture, so it's a major research effort. And it leads to much more authentic testing. It allows you to look at aspects of performance that you just cannot do with paper and pencil, such as what people understand when you speak to them, and how they speak to you. But a whole new testing methodology and a whole new view of what testing is has to be developed, and that's where the effort has to go.

5c. What suggestions do you have for how to best take advantage of these future possibilities or to overcome potential difficulties?

Response from: Margaret Honey

I believe there needs to be a Federal commitment to ensuring equity in access. Without this commitment we will continue to have situations in which wealthy suburban school districts have every advantage over rural and urban schools. At the local and state level we need to substantially rethink learning and teaching. The curriculum, across the board, needs to be focused on research, inquiry and interpretation -- not memorization. Students and teachers need longer class periods, and teachers need opportunities to learn and collaborate just like their students. We need a collection of exemplars that demonstrate for others schools and districts that have used communications technologies for innovative teaching, learning and assessment.

5c. What suggestions do you have for how to best take advantage of these future possibilities or to overcome potential difficulties?

Response from: Jeffrey Munks

I think the single most important step in accelerating the process of development and implementation is to encourage ongoing dialogue across the four sectors that for far too long have been talking only to themselves - education, government, business, and the community. From the grass-roots to the national level, these four sectors all contribute to the mix and all should be at the table. Working together, they will see areas of common concern and each will discover resources in the others that were there all along but were never brought to bear in ways that could help in another sector. Examples of the benefits of cross sector dialogue can be seen in the work being done on National Standards for Foreign Language Instruction, K-12, (cross sector advisory board) and in the construct of the National Advisory Board of the National Foreign Language Resource Center at Ohio State University. Also, Dr. Ron Walton and Dr. Richard Brecht of the National Foreign Language Center have done significant work in outlining the value of cross sector collaboration in support of the kinds of technology issues described in your survey. I would suggest that they might be able to provide excellent advice and counsel as your work continues.

Finally, and perhaps most ambitiously, an effort to catalog, assess, and coordinate the various technology experiments, applications, etc. that are going on all around us would be worthwhile. It would undoubtedly uncover unnecessary duplications, cost saving opportunities, and could lead to the acceleration of migration of successful trials and programs. Such an effort could be led by any number of organizations but would logically (I think) be led by the Office of Education.

5c. What suggestions do you have for how to best take advantage of these future possibilities or to overcome potential difficulties?

Response from: David Niguidula

We should begin by taking advantage of the schools that are thinking about reform, and designing new forms of assessment, and asking them how technology would be useful. The educational issues need to come first; when a group thinking hard about the educational issues come up with new necessary structures for communicating and interacting with information, then the technology will have a clear role.

The use of technology, by itself, is not the goal; the goal is to examine the forms of assessment that we want to create and then determine how technology can help.

Above all, schools need coaching or some practice in determining what is possible. We need to help all involved in schooling understand that education and schooling are not totally identical. If our goal is an educated populace, then we need structures that allow students to show what they know and can do. By beginning the conversation here, rather than on the technology, we will see more effective use of the technology, since it will fulfill a need, rather than be a solution searching for a question.

Special Issues Analysis Center

A Technical Support Center for the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education

Operated by:

Development Associates, Inc.

1730 North Lynn Street, Arlington, VA 22209-2023

Tel: (703) 276-0677 Fax: (703) 276-0432

and its subcontractor:

Westat, Inc.

1650 Research Blvd., Rockville, MD 20850-3129

BEST COPY AVAILABLE